

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07K 14/00	A2	(11) International Publication Number: WO 98/24810 (43) International Publication Date: 11 June 1998 (11.06.98)
(21) International Application Number: PCT/EP97/06956 (22) International Filing Date: 3 December 1997 (03.12.97) (30) Priority Data: 9625283.8 4 December 1996 (04.12.96) GB (71) Applicant (for all designated States except US): JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B-9032 Wondelgem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Travessera de les Corts, 171/702a, E-08028 Barcelona (ES). DERAEMYAEKER, Marc [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VERHASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F-34000 Montpellier (FR). MAERTENS, Luc, Jacques, Simon [BE/BE]; Vier Uitersten 26, B-8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30,		B-2340 Beerse (BE). VANDEKERCKHOVE, Joel, Stefaan [BE/BE]; Rode Boukendreef 27, B-8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE). (74) Agent: BALDOCK, Sharon, Claire; Boulton Wade Tennant, 27 Fumival Street, London EC4A 1PQ (GB). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS		
(57) Abstract Vertebrate protein homologues of UNC-53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

VERTEBRATE HOMOLOGUES OF UNC-53
PROTEIN OF C. ELEGANS

The present invention relates to vertebrate
5 homologues of UNC-53 protein of C. elegans and cDNA
sequences coding for said homologues or functional
equivalents thereof. The invention also relates to
processes for identifying compounds which control cell
behaviour, compounds identified and pharmaceutical
10 compositions containing them in addition to processes
and assays for identifying disease states in which
said gene or protein is dysfunctional.

The control of cell motility, cell shape and
directionality of cell outgrowth of axones or other
15 cell outgrowths is an essential feature in the
morphogenesis and function of both unicellular and
multicellular organisms. The control of these
processes is disturbed in a variety of disease states
in which, for example, the Receptor Tyrosine kinase
20 (RTK) signal transduction pathways, or the like, or
their downstream intra-cellular pathways (which are
shared with other extra-cellular receptors, including
cell adhesion molecules like N-CAMS and integrins) are
overstimulated.

25 Some cell surface proteins and extra-cellular
molecules controlling the directionality and potential
of cell migration have been identified, although the
processes involved are not generally understood.

It is generally considered that a long-range
30 migration of a cell process (also known as a growth
cone extension) is a stepwise event, whereby prior to
and after each extension there is the formation of a
structure at the leading edge of the cell which senses
signals in the environment instructing the cell to
35 either stabilise a cell process extending in a

- 2 -

preferred direction, or to cause a lamellipodium to extend a process in a given direction. Localised stabilisation of the actin cytoskeleton and association with plus end regions of microtubules is a general cell biological process underlying the choice of directional extension. Microtubule binding directing these processes has not previously been identified. The present inventors have surprisingly found that UNC-53 protein of *C. elegans* and vertebrate homologues thereof is involved in binding of microtubules and particularly of plus end regions of microtubules.

A gene from the free-living nematode *Caenorhabditis elegans* designated "unc-53" has been previously identified and cloned (Abstract, International *C. elegans* Meeting, June 1-5 1991, Madison, Wisconsin, 58, Bogaert and Goh). The present inventors previously identified UNC-53 protein as a signal transducer or signal integrator controlling the directionality of cell migration and/or cell shape in *C. elegans* (WO 96/38555). Increased UNC-53 protein activity was found to be proportional to cell process extensions in the correct direction of cell migration. The unc-53 gene was found to encode a signal transduction molecule that transduces a signal from an RTK such as, for example, via the adaptor protein SEM-5/GRB-2, to the machinery controlling directional growth cone extension or stabilisation, in a highly dosage - dependent fashion.

Genetic and experimental analysis of *C. elegans* UNC-53 mutants showed that mutations in the unc-53 gene do not affect the general ability of cells to migrate but rather affect the ability of cells to migrate under specific antero-posterior cues. Reduction of UNC-53 activity leads to loss of

- 3 -

direction and reduction of growth cone extension as indicated by the directionality of random extension cycles observed in excretory canal growth cones in UNC-53 mutants.

5 The function of UNC-53 is highly sensitive to its dosage or activity. Reduction of function leads to proportional reduction of migration to the specific signal while increased expression, using transgenic expression of UNC-53 in muscle cells, leads to
10 increased directional migration. The data lead to the conclusion that UNC-53 functions as an integrator of a directional signal in the organism whereby reception of signals leads to growth cone extension in the correct direction.

15 Certain alleles of UNC-53 enhance the sex myoblast migration defect of SEM-5 C. elegans mutants in a receptor tyrosine kinase signal transduction pathway (Stern et al 1993 mol. Biol. cell, 4, 1175-1188). While the genetics suggests that UNC-53 and
20 SEM-5 cooperate to regulate sex myoblast migration, genetic experiments do not permit a conclusion that this is the result of a direct molecular interaction. The inventors previously identified a potential sem-5/GRB-2 binding site and showed in two types of
25 biochemical experiments that UNC-53 physically interacts with SEM-5. The present inventors conclude that UNC-53 encodes a signal transduction molecule that transduces extracellular signals for directional migration via the adapter protein SEM-5/GRB-2 to the
30 machinery controlling directional growth cone extension or stabilization.

 Several lines of evidence indicate that UNC-53 might act as an adapter linking extracellular signals to the actin cytoskeleton. Firstly, UNC-53 has shown
35 homology to cortical actin binding proteins and that

- 4 -

it is capable of binding F-actin in vitro. In addition expression of UNC-53 in mammalian cells leads to changes in the F-actin cytoskeleton. Very low levels of UNC-53 expression increase the number of
5 filopodia and actin microspikes protruding from the cell surface. Cells expressing UNC-53 also exhibit increased neurite extension and increased cell motility. UNC-53 thus also acts as an activator of migration.

10 Considering all available data the following possible mechanisms of action of UNC-53 can be formulated.

The choice and activation of directional growth cone extension can be accounted for by local
15 activation of UNC-53 via a SEM-5/GRB2 complex to a receptor (eg receptor tyrosine kinase signal) which reads a localized or directional signal. Changes in growth cone steering are preceded by the formation of a localized actin patch in the area of the growth cone
20 receiving the highest signal (Bentley and O'Connor et al. Curr. Op. NeuroBiol. 1994, vol 4, 43-48). UNC-53 might be directly involved in forming these actin patches through its own actin binding or cross-linking properties. Alternatively activated UNC-53
25 may (eg via its nucleotide binding domain) transduce a signal to as yet unidentified effectors. For example, activation of the small GTP-binding protein cdc42 or a related protein leads to formation of small actin patches as well as the formation of small filopodia.
30 The unc-53 pathway may be upstream of cdc42 or both signal transducers might share downstream pathways.

The present inventors thus decided to investigate if a similar protein was present in higher organisms such as vertebrates.

35 The present inventors describe the identification

- 5 -

of a family of genes in vertebrates, and particularly in man and mouse with extensive structural homology to UNC-53. The present inventors have surprisingly found that the nucleotide domains of UNC-53 from C. elegans and UNC-53 from vertebrates similarly activate motility, establishing functional equivalence. Furthermore these domains are shown to be capable of transforming NIH3T3 cells in vitro. The inventors also found changes in RNA transcripts in transformed cell lines compared to normal human tissues suggesting a role for UNC-53 in cell differentiation, morphogenesis and disease. Furthermore, in vitro assays and transgenic models are also described that identify pharmacological modulators of UNC-53 activity and assays to identify proteins interacting with UNC-53.

According to a first aspect of the present invention, there is provided a vertebrate protein homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein homologue comprises an amino acid sequence having a statistically significant homology to the UNC-53 protein of C. elegans as illustrated in figure 2. According to the present invention a derivative should be taken to mean mutational derivatives, fusions, internal deletions, splice variants and muteins.

There is also provided according to a second aspect of the present invention a vertebrate protein homologue of UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence homology blocks A, B, C, D or E as illustrated in Figure 9a, or block F in Figure 12a or a sequence having a statistically significant homology therewith.

- 6 -

Preferably, said vertebrate homologue is a human protein or a mouse protein.

According to a further aspect of the invention there is provided a vertebrate protein homologue of an
5 UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E or F which differ from those blocks of Figure 9a and Figure 12a to a significant extent only in conservative amino acid changes. In an
10 even further aspect of the invention there is provided a vertebrate protein having an amino acid sequence encoded by the nucleotide sequence from position 1 to position 6013 as illustrated in Figure 9b. There is also provided a vertebrate protein having an amino
15 acid sequence encoded by the nucleotide sequence illustrated in Figure 11d, or a functional equivalent derivative, or bioprecursor of said homologue.

According to a further aspect of the present invention there is provided a vertebrate protein
20 having an amino acid sequence corresponding to the prosite signatures as illustrated in Figure 28 for each of said homology blocks as defined above. Advantageously the prosite signatures can be used to identify a protein having a statistically significant
25 homology to the UNC-53 protein of C. elegans. (Luethy et al 1994, Protein Science, 3, 139-146).

A further aspect of the invention comprises a vertebrate homologue according to the invention comprising an amino acid sequence as shown in figure
30 9b or 11d or an amino acid sequence which differs from the amino acid sequences shown in these figures to a significant extent only in one or more conservative amino acid changes.

In a further aspect of the present invention
35 there is also provided a nucleic acid molecule, which

- 7 -

is preferably DNA, and which encodes a vertebrate homologue of UNC-53 protein of C. elegans, or a functional equivalent derivative, fragment or bioprecursor of said homologue according to the invention. Preferably, the cDNA comprises a sequence of nucleotides encoding an amino acid sequence as illustrated in figures 9b or 11d or an amino acid which differs from the sequences shown in these figures to a significant only in one or more conservative amino acid changes. Preferably the DNA is cDNA, which cDNA comprises at least from position 1 to 6013 of the sequence shown in Figure 9b. Alternatively the cDNA may comprise the sequence illustrated in Figure 11d. Also provided by the present invention is a nucleic acid sequence capable of hybridising to the nucleic acid or DNA sequences according to the invention under high stringency conditions, which conditions are well known to those skilled in the art.

20 The cDNA according to the invention may be included in an expression vector which may itself be used to transform or transfect a host cell, which cell may be bacterial or eukaryotic in origin including such as, for example an animal or plant cell a fungal cell or an insect cell. Thus, advantageously, once the cDNA corresponding to the genome of the vertebrate homologue of UNC-53 of C. elegans is synthesised, using for example, reverse transcriptase or the like, a range of cells, tissues or organisms may be transfected following incorporation of the selected cDNA clone into an appropriate expression vector. The expression vector according to the invention may comprise a promoter of C. elegans or one of human mouse or viral origin and optionally a sequence encoding a reporter molecule, such as, for example,

- 8 -

green fluorescent protein.

The present invention, therefore, also further comprises a transgenic cell, tissue or organism comprising a transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment derivative or bioprecursor of said homologue. The term "transgene capable of expressing a vertebrate homologue of UNC-53 protein of C. elegans" as used herein means a suitable nucleic acid sequence which leads to the expression of a vertebrate homologue of UNC-53 protein of C. elegans having the same function and/or activity. The transgene may include, for example, genomic nucleic acid isolated from the appropriate vertebrate or synthetic nucleic acid including cDNA. The term "transgenic organism, tissue or cell, as used herein means any suitable organism and/or part of an organism, tissue or cell, that contains exogenous nucleic acid either stably integrated in the genome or in an extrachromosomal state.

Preferably the transgenic cell comprises any of, a COS cell, HepG2 cell, MCF-7 or N4 neuroblastoma cell or a NIH3T3 cell or a colorectal or carcinoma cell or a human derived cell such as a fibroblast or the like. The transgenic organism may be an insect, a non-human animal or a plant and preferably C. elegans or a related nematode. Preferably, the transgene comprises the nucleic acid sequence encoding the vertebrate homologue or a functional fragment of said gene according to the invention as described above. The transgene preferably comprises an expression vector according to the invention.

The term "functional fragment" as used herein should be taken to mean a fragment of the gene coding for the vertebrate homologue of the UNC-53 protein of

- 9 -

C. elegans or a functional equivalent or derivative or bioprecursor of said protein. For example, the gene may comprise deletions or mutations but may still encode a functional vertebrate homologue of UNC-53 protein.

Further provided by the present invention is a method of producing a mutant vertebrate non-human organism or cell having a mutation in the wild-type gene coding for the vertebrate homologue of UNC-53 protein, which mutation affects cell behaviour or the regulation of cell motility or the shape or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes located thereon, which method comprises inducing a mutation in the vertebrate homologue of UNC-53 protein in said organism or cell. These mutant organisms or cells may be used in a screen to identify the effects of compounds on these cell functions.

The vertebrate homologue of UNC-53 protein of C. elegans or the cDNA or genomic DNA encoding it or a functional equivalent, derivative, fragment or bioprecursor of said homologue, may advantageously be used as a medicament, or in the preparation of a medicament to promote neuronal regeneration, revascularisation or wound healing or the treatment of chronic neurodegenerative disorders or acute traumatic injuries or fibrotic disease or physiological events requiring the polarity of cells or epithelia. The present inventors have also found that the vertebrate homologue of UNC-53 protein plays a role in a transformed state of cells. Accordingly, the vertebrate homologue, dominant positive or negative mutants thereof, or inhibitors thereof may advantageously be used to induce or alleviate contact inhibition in a cell or in preventing cancer

- 10 -

development. Typically, the above medical conditions may be treated in mammals and more preferably humans by either a homologue of UNC-53 protein or alternatively by a nucleic acid coding for such a protein. Alternatively an antisense oligonucleotide to said UNC-53 homologue may be used to prevent its expression. Examples of other nucleic acid sequences which may be used include 3' untranslated regions of mRNA which could be used to prevent transcription of the genomic sequence encoding for the vertebrate homologue of UNC-53 protein.

The vertebrate homologue of UNC-53 protein or a functional equivalent, fragment or bioprecursor of said protein may be incorporated into a pharmaceutically acceptable composition together with a suitable carrier, diluent or excipient therefor. The pharmaceutical composition may advantageously comprise, additionally or alternatively, the nucleic acid sequence according to the invention as defined above.

The present invention also provides for a method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, transformation, cell shape or motility or the direction of cell migration or microtubule plus end stability or function and localisation of protein complexes thereon which method comprises contacting said compound with a transgenic cell according to the invention and screening for a phenotypic change in said cell. Preferably the method can determine whether the compound comprises an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which pathway said vertebrate homologue of UNC-53 protein, or a functional equivalent, derivative, fragment or bioprecursor of

- 11 -

said vertebrate homologue is a component or whether said compound is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell. The present invention also provides a method to determine that the protein in said signal transduction pathway is a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

10 Preferably, the phenotypic change to be screened comprises a change in cell shape or a change in cell motility. Where a transgenic cell is used in accordance with one embodiment of the method of the invention, an N4 neuroblastoma cell may be used and in
15 such an embodiment the phenotypic change to be screened may be the length of neurite growth or changes in filipodia outgrowth or alternatively changes in ruffling behaviour or cell adhesion or any change in microtubule cytoskeleton or any change in
20 localisation of proteins on plus end regions of microtubules or any change in cell death such as in apoptosis. In an alternative embodiment of the method of the invention, the transgenic cell may comprise an MCF-7 breast cancer cell. Typically in such an
25 embodiment the phenotypic change to be screened comprises the extent of phagokinesis or filipodia formation. In an alternative embodiment of this aspect of the invention, the transgenic cell may comprise an NIH3T3 cell. Typically in such an
30 embodiment the phenotypic change to be screened comprises loss of contact inhibition of foci formation. The method according to the invention, may also utilise a mutant cell or mutant organism according to the invention as described above, where
35 the mutant cell is capable of growing in tissue

- 12 -

culture or in vivo and either of which cell or organism has a mutation in the wild-type unc-53 gene.

In accordance with the present invention, a "phenotypic change", may be any phenotype resulting from changes at any suitable point in the life cycle of the cell, tissue or organism defined above, which change can be attributed to the expression of the transgene such as for example, growth, viability, morphology, behaviour, movement, cell migration or cell process or growth cone extension of cells and includes changes in body shape, locomotion, chemotaxis, contact inhibition, mating behaviour or the like. The phenotypic change may preferably be monitored directly by visual inspection of the cell as a whole or particularly by monitoring the F-actin cytoskeleton microtubule network and plus end stability of microtubules or proteins thereon or alternatively by for example measuring indicators of viability including endogenous or transgenically introduced histochemical markers or other reporter genes, such as for example β -galactosidase or green fluorescent protein.

A compound which is identifiable by the method according to the invention as described above, as an enhancer of the processes identified above such as the regulation of cell shape or motility or the direction of cell migration may be used as a medicament, or alternatively in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease. Examples of promoting neuronal regeneration include, for example, peripheral nerve regeneration after trauma and spinal cord trauma.

Where a compound is identified in accordance with

- 13 -

the method described above as being an inhibitor of the regulation of cell shape etc., the compound may be used as a medicament, or in the preparation of a medicament, for substantially alleviating spread of disease inducing cells, such as in spread of cancer, or the like in metastasis or in alleviating loss of contact inhibition. Advantageously, any of the compounds which may have been identified as an inhibitor or an enhancer in accordance with the method as described above, may also be included in a pharmaceutical composition comprising the respective compound and a pharmaceutically acceptable carrier, diluent or excipient therefor.

The particular mechanism of action of a compound identified as either an inhibitor or an enhancer of the cell motility shape, growth or direction of cell migration or microtubule association or to the plus end region thereof is not limiting. Preferably the compound acts as an inhibitor or enhancer of a signal transduction pathway. The compound may also act on a parallel pathway or directly on the vertebrate homologue of UNC-53 protein of C. elegans. For example, the method of action of the compound may include direct interaction with the vertebrate homologue of UNC-53 protein, interaction with processes for regulating phosphorylation or dephosphorylation of the vertebrate homologue of UNC-53 or with processes regulating activity of an unc-53 gene or with processes for post-transcriptional or post-translational modification or the like.

Preferably the compound is identified by the method according to the invention as an inhibitor or an enhancer, by utilising differences of phenotype of the cell, tissue or organism, which are visible to the eye. Alternatively indicators of viability including

- 14 -

endogenous or transgenically introduced histochemical markers or a reporter gene may be used.

According to a further aspect of the invention there is also provided a transgenic cell or tissue
5 culture which has been constructed to comprise a promoter sequence of a gene coding for a vertebrate homologue of UNC-53 of C. elegans or a functional equivalent, derivative fragment, or bioprecursor of said homologue operably linked to a nucleic acid
10 sequence encoding a reporter molecule. Preferably, the reporter sequence encoding the reporter molecule which comprises a detectable protein, for example one which may be monitored by eye inspection such as antibiotic resistance, β -galactosidase or a molecule
15 detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also provides a method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a
20 vertebrate homologue of UNC-53 protein in C. elegans, or a functional equivalent, derivative fragment or bioprecursor of said homologue, which method comprises the steps of:

- 25 (a) contacting said compound with a transgenic cell according to the invention as described above,
- (b) monitoring the level of said reporter molecule and comparing results obtained from this monitoring step with a control comprising a
30 transgenic cell having the promoter sequence of a gene coding for a vertebrate homologue of UNC-53 protein, or a functional fragment of said homologue and the reporter molecule, in the absence of the compound.

35 In one embodiment of the method according to this

- 15 -

aspect of the invention the reporter molecule may comprise messenger RNA.

5 A compound identified as an enhancer of transcription of the gene coding for the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor of said homologue may also be used as a medicament, or in the preparation of a medicament, for promoting neuronal regeneration, revascularisation or wound
10 healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease. Furthermore, such compounds may be included in a pharmaceutical composition including a pharmaceutically acceptable carrier, diluent or
15 excipient therefor. Any compounds identified as inhibitors of transcription may, advantageously, be used in alleviating the spread of disease inducing cells such as cancers or metastasis or loss of contact inhibition.

20 The present invention also provides a kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell growth, transformation, cell motility or shape or the direction of cell migration which kit comprises at
25 least one transgenic or mutant cell or transgenic or mutant non-human organism according to the invention as described above and a plurality of wild-type cells or one organism of the same type, or a cell line or tissue culture and means for contacting said compound
30 with said cell or organism.

Also provided by the present invention is a kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of UNC-53 protein of C. elegans
35 or a functional equivalent, derivative or fragment

- 16 -

thereof, which kit comprises at least one transgenic cell or cells according to the invention and means for contacting said compounds with said cells.

For the purposes of the present invention, the
5 term "gene coding for a vertebrate homologue of UNC-53
or a functional fragment of said homologue" includes
the nucleic acid sequence shown in Figures 9b or 11d
or a fragment thereof, including the differentially
spliced isoforms and transcriptional starts of the
10 nucleic acid sequence and which sequence encodes a
vertebrate homologue of UNC-53 protein or a functional
equivalent, derivative, fragment or bioprecursor of
the protein.

The present invention also provides methods of
15 identifying genes of vertebrates or fragments of said
genes, which encode proteins which are active in the
signal transduction pathway of which the vertebrate
homologue of UNC-53 is a component. A preferred
method comprises hybridizing to an appropriate cDNA
20 library a nucleotide sequence, as defined herein, or a
fragment thereof under appropriate conditions of
stringency in order to identify genes having
statistically significant homology with the cDNA
clones of any one of the cDNA sequences according to
25 the invention described above.

Furthermore, there is also provided by the
present invention a method of identifying a protein
which is active in the signal transduction pathway of
a cell of which a vertebrate homologue of UNC-53
30 protein of C. elegans or a functional equivalent,
fragment or bioprecursor of said vertebrate homologue
is a component. According to this aspect of the
invention, the method comprises;

(a) contacting an extract of said cell with an
35 antibody to the vertebrate homologue of UNC-53

- 17 -

protein or a functional equivalent, fragment or bioprecursor of said protein,

(b) identifying the antibody/vertebrate homologue of UNC-53 complex, and

5 (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the antibody.

The vertebrate homologue of UNC-53 protein, therefore may bind regions of other proteins involved
10 in the signal transduction pathway. It is also possible to sequentially identify a whole range of proteins involved in the signal transduction pathway.

Antibodies to the vertebrate homologue of UNC-53 protein may be produced according to known techniques
15 as would be known to those skilled in the art. For example, polyclonal antibodies may be prepared by inoculating a host animal, such as a mouse, with a protein or epitope of a protein according to the invention and recovering immune serum.

20 This aspect of the invention further comprises a method of identifying a further protein or proteins which are active in the signal transduction pathway of a cell of which UNC-53 is a component which method comprises:

25 (a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein in the method as described above,
(b) contacting a cell extract with the antibody,
(c) identifying the antibody/protein complex,
30 (d) analysing the complex to identify any further protein bound to the first protein other than the antibody, and
(e) optionally repeating steps (a) to (d) to identify further proteins in the pathway.

35 According to this aspect of the present

- 18 -

invention, the antibody starts the process by binding to the vertebrate homologue of UNC-53 protein or a functional equivalent thereof in the signal transduction pathway. Any other proteins found
5 complexed to the bound antibody or UNC-53 protein can then be used to identify further interacting proteins involved in the pathway.

It may also be possible to identify proteins involved in the signal transduction pathway of a cell
10 of which the vertebrate homologue of UNC-53 or a functional equivalent derivative or bioprecursor thereof is a component by using a vertebrate homologue of UNC-53 protein of C. elegans. According to this aspect of the invention the method comprises:

- 15 (a) contacting an extract of the cell with the vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue,
- (b) identifying the vertebrate homologue of
20 UNC-53 protein/protein complex formed and
- (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein

25 This method can also advantageously be used to identify further proteins in a signal transduction pathway of a cell by contacting an extract of the cell used as described above, with any protein identified from step (c) above not being a vertebrate homologue
30 of UNC-53 protein and repeating steps (b) and (c).

Other methods which may be used for identifying proteins in a signal transduction pathway of a cell may comprise for example a western blot overlay method which method is well known to those skilled in the
35 art. Cell extracts are run on gels to separate out

protein and subsequently blotted onto a nylon membrane. These membranes may then be incubated, for example in a medium containing a vertebrate homologue of UNC-53 having a label attached thereto such as a biotin or radiolabel and any protein conjugates visualised with for example a streptavidin or alkaline phosphatase conjugated antibody.

The present invention also advantageously provides a process for the preparation of binding antibodies which recognise proteins or fragments thereof involved in the rate and direction of cell migration or the control of cell growth or shape, for the above methods.

The monoclonal antibody for binding to the appropriate vertebrate homologue of UNC-53 (or its functional equivalent) may be prepared by known techniques as described by Kohler R. and Milstein C., (1975) Nature 256, 495 to 497.

Another method which may be used to identify proteins involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent or derivative or bioprecursor is a component involves investigating protein-protein interactions using the two-hybrid vector method. This method is well known to those skilled in the art and which was first developed in yeast by Chien et al (1991). This technique is based on functional reconstitution in vivo of a transcription factor which activates a reporter gene. More particularly the technique comprises providing an appropriate host cell with a DNA construct comprising a reporter gene under the control of a promoter regulated by a transcription factor having a DNA binding domain and an activating domain, expressing in the host cell a first hybrid DNA

- 20 -

sequence encoding a first fusion of a fragment or all of a nucleic acid sequence according to the invention and either said DNA binding domain or said activating domain of the transcription factor, expressing in the host at least one second hybrid DNA sequence, such as a library or the like, encoding putative binding proteins to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion; detecting any binding of the proteins to be investigated with a protein according to the invention by detecting for the presence of any reporter gene product in the host cell; optionally isolating second hybrid DNA sequences encoding the binding protein.

15 An example of such a technique utilises the GAL4 protein in yeast. GAL4 is a transcriptional activator of galactose metabolism in yeast and has a separate domain for binding to activators upstream of the galactose metabolising genes as well as a protein binding domain. Nucleotide vectors may be constructed, one of which comprises the nucleotide residues encoding the DNA binding domain of GAL4. These binding domain residues may be fused to a known protein encoding sequence, such as for example a sequence coding for the vertebrate homologue of UNC-53. The other vector comprises the residues encoding the protein binding domain of GAL4. These residues are fused to residues encoding a test protein, preferably from the signal transduction pathway of the vertebrate in question. Any interaction between the vertebrate homologue of UNC-53 protein and the protein to be tested leads to transcriptional activation of a reporter molecule in a GAL-4 transcription deficient yeast cell into which the vectors have been transformed. Preferably, a reporter

- 21 -

molecule such as β -galactosidase is activated upon restoration of transcription of the yeast galactose metabolism genes. This method enables any interactions between proteins involved in the signal transduction pathway or a parallel or redundant pathway to be investigated.

Any proteins identified in the signal transduction pathway of the cell, which may be for example a mammalian cell, may also be included in a pharmaceutical composition together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

The present invention also provides a process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, or derivative of the protein, which process comprises culturing the cells transformed or transfected with a cDNA expression vector having any of the cDNA sequences according to the invention as described above, and recovering the expressed vertebrate homologue of UNC-53 protein. The cell may advantageously be a bacterial, animal, insect or plant cell.

A particularly preferred process for producing a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or fragment of said homologue comprises using insect cells. Accordingly, the invention provides a process for producing a vertebrate homologue of UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of the UNC-53 protein, which process comprises culturing an insect cell transfected with a recombinant Baculovirus vector, said vector comprising a nucleotide vector encoding the vertebrate homologue of UNC-53 protein or a functional equivalent, fragment

- 22 -

or bioprecursor thereof downstream of the Baculovirus polyhedrin promoter and recovering the expressed protein. Advantageously, this method produces large amounts of protein for recovery. The insect cell may
5 be from for example Spodoptera frugiperda or Drosophila Melanogaster.

In accordance with the present invention, a defined nucleic acid sequence includes not only the identical nucleic acid but also any minor base
10 variations from the natural nucleic acid sequence including in particular, substitutions in bases which result in a synonymous codon (a different codon specifying the same amino acid), due to the degenerate code in conservative amino acid substitution. The
15 term "nucleic acid sequence" also includes the complimentary sequence to any single stranded sequence given which includes the definition above regarding base variations.

Furthermore, a defined protein, polypeptide or
20 amino acid sequence according to the invention, includes not only the identical amino acid sequence but also minor amino acid variations from the natural amino acid sequence including conservative amino acid replacements (a replacement by an amino acid that is
25 related in its side chains). Also included are amino acid sequences which vary from the natural amino acid but result in a polypeptide which is immunologically identical or similar to the polypeptide encoded by the naturally occurring sequence. Such polypeptides may
30 be encoded by a corresponding nucleic acid sequence.

A further aspect of the invention provides a nucleic acid sequence of at least 15 nucleotides of a nucleic acid according to the invention and preferably from 15 to 50 nucleotides.

35 These sequences may, advantageously be used as

- 23 -

probes or primers to initiate replication or the like. Such nucleic acid sequences may be produced according to techniques well known in the art, such as by recombinant or synthetic means. They may also be used
5 in diagnostic kits or the like for detecting for the presence of a nucleic acid according to the invention. These tests generally comprise contacting the probe with a sample under hybridising conditions and detecting for the presence of any duplex formation
10 between the probe and any nucleic acid in the sample. Nucleic acid sequences according to the invention may also be produced using recombinant or synthetic means such as described in Sambrook et al (Molecular Cloning: A Laboratory Manual, 1989). Advantageously,
15 human allelic variants or polymorphisms of the DNA according to the invention may be identified by, for example, probing DNA libraries from a range of individuals for example from different populations. Furthermore, nucleic acids and probes according to the
20 invention may be used to sequence genomic DNA from patients using techniques well known in the art, such as the Sanger Dideoxy chain termination method, which may advantageously ascertain any predisposition of a patient to certain proliferative disorders.

25 A method of detecting whether a compound is an inhibitor or an enhancer of expression of a vertebrate homologue of UNC-53 of C. elegans, or a functional equivalent, derivative or fragment of said vertebrate homologue is also provided which method comprises
30 contacting a cell expressing said homologue with said compound and monitoring for a phenotypic change compared to a control cell which has not been contacted with said compound.

35 Preferably the cell is a transgenic cell as described above. Alternatively the cell may have

- 24 -

undergone loss of contact inhibition.

The present method also provides for determining whether said compound is an inhibitor of expression of said vertebrate homologue. In one embodiment the
5 compound to be tested comprises a nucleic acid.

Preferably said nucleic acid sequence comprises an antisense DNA sequence or a mRNA sequence.

Preferably said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said
10 vertebrate homologue.

Alternatively, the compound to be tested may be a protein. Preferably, said protein comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue and
15 preferably comprises a protein identified by the methods as described herein.

The present invention also provides a pharmaceutical composition comprising a compound, for example an antisense nucleic acid identified according
20 to the above described method together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

A nucleic acid sequence or protein identified according to this aspect of the invention may be used
25 as a medicament, or in the preparation of a medicament, for treating loss of contact inhibition or cancer which is mediated by a vertebrate homologue of UNC-53 protein or a functional equivalent, fragment, derivative or bioprecursor of said homologue.

Further provided by the invention is a nucleic acid as defined above for use in preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of UNC-53 protein of
30 C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said homologue.
35

- 25 -

According to a further aspect of the invention there is provided a plasmid pCB201 deposited under LMBP Accession No. LMBP 3594 and a MCF-7 and a NIH/3T3 cell line transfected with plasmid pCB201 deposited
5 under LMBP Accession Nos. LMBP 1601 CB and LMBP 1603 CB respectively. Further provided by the invention is phage lambda 3b coding for Hu-UNC-53/1 and deposited under Accession No. LMBP 1604CB (or 3595). Also provided are plasmids pLM1 deposited under Accession
10 No. LMBP 3762, pLM4 (LMBP 3763), pEGFP72 (LMBP 3764) and pCB501 (LMBP 3765). Further provided is a Bac clone comprising a fragment of hu-unc-53/2 gene (LMBP 3773) and a worm strain comprising a chimeric C.elegans human unc53 gene deposited under LMBP
15 Accession No. LMBP-1663CB.

Further provided by the invention is an assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract
20 thereof with an antibody to said vertebrate homologue, or a functional equivalent, derivative or bioprecursor thereof, which antibody is fused to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

25 Preferably the reporter molecule is an antibody conjugated to for example a flurophore such as fluorescein or alternatively to an enzyme such as strepavidin.

There is also provided a method for detecting for
30 expression of a gene coding for a vertebrate homologue of UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or
35 corresponding to said vertebrate homologue or a

- 26 -

functional equivalent, fragment or bioprecursor thereof with a cell extract, which probe is linked to a reporter and analysing for the presence of said reporter.

5 Preferably the probe is a complementary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

10 Preferably the complimentary sequence is a 3' or 5' untranslated region of said mRNA. Preferably said reporter may be a dig label, a fluorophore, a hapten or a radiolabel.

 Alternatively said probe comprises an antibody
15 specific for said vertebrate homologue of said UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor therefor.

 Preferably the reporter is an antibody conjugated to for example a fluorophore such as fluorscein or
20 alternatively an enzyme such as streptavidin.

 As described above UNC-53 protein of C.elegans has been found to localise to microtubule and particularly to microtubule (+) ends. Therefore, there is provided by a further aspect of the present
25 invention a method of determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to any of claims to 1 to 9 to microtubules or plus end regions thereof, which method comprises (a) contacting
30 said compound with a transgenic cell, tissue or organism expressing UNC-53 protein or said vertebrate homologue and which protein is operably linked to a reporter molecule (b) screening for the localisation of said reporter molecule as compared to a cell
35 according to step (a) which has not been contacted

- 27 -

with said compound.

A compound identifiable by the above method also forms part of the present invention. Such a compound identified as an inhibitor of localisation or
5 association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition. Further a compound identified as an enhancer of
10 association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof may be used in for example promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic neurodegenerative diseases or
15 acute traumatic injuries or fibrotic disease. These compounds may then be included in a pharmaceutical composition, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

Also provided by the present invention is a kit
20 for determining whether a compound is an inhibitor or an enhancer of association of UNC-53 or a vertebrate homologue thereof according to the invention with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing
25 UNC-53 and a reporter molecule or a host or transgenic cell according to the invention and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells. Compounds identified as
30 inhibitors or enhancers or microtubule association described above may advantageously be included in a composition and linked to unc-53 protein of C.elegans or a vertebrate homologue thereof according to the invention to target the compounds to the microtubules
35 or the plus end regions thereof. Such a composition

- 28 -

may also comprise, for example, a suitable transfecting or transformation agent.

According to a further aspect of the invention there is provided a method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to the invention, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed and which results in targeting said protein to said microtubule or a plus end region thereof. An even further aspect of the invention comprises a method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue thereof according to the invention, which method comprises a) contacting either an extract from a cell or cells expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising candidate UNC-53 modifying enzymes in the presence of an indicator of covalent modification of a protein, b) identifying any covalently modified UNC-53 protein from step a) and c) identifying said molecule involved in said modification step. Such an indicator may be ³²P.

Further provided by the invention is a method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to the invention, or which alleviates or enhances apoptosis. The method of the former comprises contacting said compound with a transgenic cell, tissue or organism according to the invention and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof. In the case of apoptosis the

- 29 -

method comprises monitoring the effect of the compound on cell death.

The invention may be more clearly understood from the following examples which are only exemplary, with
5 reference to the accompanying drawings wherein

Figure 1 illustrates the sequence of plasmid pTB72 which codes for the full length UNC-53 protein in C. elegans, deposited under LMBP Accession No. 3486.

10 Figure 2 illustrates the full-length UNC-53 protein from C. elegans.

Figure 3 is a Tblastn search of the EST division of Genbank with the ORF of the longest known Ce-UNC-53 cDNA. tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in Ce-UNC-53.

15 Figure 4 illustrates a search of the Genbank databases with part of the nucleotide binding domain of Ce-UNC-53. It does not identify statistically significant proteins except for the C. elegans cosmid containing Ce-unc-53.

20 Figure 5 illustrates a three frame translation of EST gb:R41071.

Regions of homology with Ce-Unc-53 in two different frames are underlined. The spacing between the blocks of homology is of similar size to that in
25 Ce-UNC-53. Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with Ce-UNC-53 (see alignment in fig. 12).

30 Figure 6 is a BLASTN search of the EST division of Genbank with Hu-unc-53/1 cDNA cosmid 3b.

Figure 7 is a TBLASTN search of the Genbank sequence database with the 961 amino acid ORF of cDNA 3b of hu-UNC-53/1 : hu-UNC-53/1 forms a unique pair
35 with Ce-UNC-53 (cosmid F45E10) compared to the rest of

- 30 -

the database.

Figure 8 is a diagram illustrating the length and overlap and tissue source of the different cDNA clones of the 3' end of Hu-UNC-53/1 isolated in this work.

5 Figure 8a. is a diagram illustrating the further sequence of the Hu-UNC-53/1 and overlap of constructs to obtain the further sequence.

10 Figure 8b is a diagram illustrating the 3' end of Hu-UNC53/1 and the EST clones present in the database.

Figure 9a is an annotated sequence listing of clone 3b of hu-UNC-53/1 including the EcoR1 polylinker GAATTC. The predicted Open Reading Frame of Hu-UNC-53/1 is listed below the sequence. Blocks A B C D and
15 E which are similiar to Ce-UNC-53/1, a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2 and the 3' untranslated leader sequences are marked with arrows and labelled.

Figure 9b is an annotated sequence listing of
20 Hu-UNC-53/1 available at this moment. The predicted Open Reading Frames of Hu-UNC-53/1, pLM1, pLM3, pLM4, pCB251, pLM5 and pCB201, the homology blocks A,B,C,D and E, the position of a region which is different between Hu-UNC-53/1 and Hu-UNC-53/2, the position of
25 phh14-3, pCB212, pCB210-14, phh3b, phh15, the position of the reverse primers HU53rv1, HU53rv2, HU53rv3 and HU53rv4, the position of peptides B72628 (=28/1), B72627, B72626 and B72625 are listed below the sequence.

30 Figure 10 is an annotated sequence listing of the insert of clone gbAA049124 (EST479167) of mu-UNC-53/1. The open reading frame and 3' untranslated sequence is marked with an arrow.

35 Figure 11a is an annotated sequence listing of the insert of clone gbH09036 (EST46037) of Hu-UNC-

- 31 -

53/2.

Figure 11b is a novel DNA sequence of HU-UNC-53/2 extended by RT-PCR. This DNA sequence is not present in EST-46037 and extends the ORF beyond position 1109 of Figure 11a to an ORF from position 18 to 1793.

Figure 11c summarises how the 3' and 5' extensions of hu-unc-53/2 were made.

Figure 11d compiles the sequence of hu-unc-53/2. The boxed sequences are the primer sequences used for the respective extension steps described in the experimental methods section.

Figure 11e illustrates the sequences of the extensions summarised in figure 11c.

Figure 11f illustrates the sequence information illustrating four alternative Start sites observed for hu-unc-53/2.

Figure 12 is an illustration of a Tblastn search of the EST division of Genbank with 680aa starting at the C-terminus of the alpha actinin domain of hu-unc-53/2.

Figure 12a is an illustration of an amino acid alignment of the available sequence of C.elegans unc-53 and hu-unc-53/1 and hu-unc-53/2.

12b. is an illustration of similarity plots for Ce-unc-53 and hu-unc-53/1 (top) and for hu-unc-53/1 and hu-unc-53/2.

Figure 13 is an annotated sequence listing of expression vector pCB201 containing homology block E from Hu-UNC-53/1 cloned in a pcDNA3.1-HIS expression vector. The HIS and T7-tags, PCR primer used to modify hu-UNC-53/1 and ORF are marked.

Figure 14 is a diagram showing the alignment of the homologous regions of hu-UNC-53/1 and mu-UNC-53/1.

Figure 15 is an annotated sequence listing of expression vector pCDU3 containing part of Ce-UNC-53/1

- 32 -

cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

5 Figure 16 is an annotated sequence listing of expression vector pCDU4 containing part of Ce-UNC-53/1 cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

10 Figure 17 is an annotated sequence listing of expression vector pCDU2 containing part of Ce-UNC-53/1 cloned in expression vector pcDNA3.1. The upper ORF starts in the vector polylinker. The lower ORF starts at the first Methionine and is part of Ce-UNC-53/1.

15 Figure 18 illustrates MCF-7 cells transfected with pCB201 (upper) compared to mock transfected MCF-7 cells (phase contrast image). The control cells are spread out on the tissue culture plastic and exhibiting few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths per cell (arrowheads).

Figure 19 is a phase contrast image of MCF-7 cells, transfected with pcDNA3.1 (19a), pCDU4 (19b), pCDU3 (19c), pCDU2 (19d) and pTB72 (19e).

25 Figure 20 is an F-actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

30 Figure 21 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.1 (21a), pCDU4 (21b), pCDU3(21c), pCDU2 (21d) and pTB72 (21e).

35 Figure 22 is a phase contrast image of N4 neuroblastoma cells transfected with pcDNA3.1 (22a), pCDU4 (22b), pCDU3 (22c), pCDU2 (22d) and pTB72 (22e).

- 33 -

Figure 23 is an F-actin pattern Phalloidin (visualised with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pcDNA3.1 (23a), pCDU4 (23b), pCDU3 (23c), pCDU2 (23d) and pTB72 (23e).

5 Figure 24 illustrates phase contrast images of small (top), medium (middle) and large foci (bottom) induced in a monolayer of NIH3T3 cells by transfection with pCB201.

10 Figure 25(c) illustrates human metaphase chromosomes probed with a probe 1p34 and figures 25a and 25b indicating the chromosomal location of hu-UNC-53/1 in 1q31. Essentially the same techniques were used to assign the gene hu-unc-53/2 to chromosome locus 11p15 (25d and e) as illustrated in micrograph 15 25f.

 The ideograms 25a and 25d are from the International System for Human Cytogenic Nomenclature 1985. The ideograms 25b and 25e in which the relative band positions and arm ratios were derived from actual 20 chromosome measurements is from Cytogenet Cell Genet 65:206-219 (1994).

 Figure 26 is an expression pattern of HU-Unc53/1 and HU-Unc532 in normal human tissues and cancer cell lines.

25 Figure 27 is a sequence map of Plasmid pNP3.

 Figure 28 is an exemplary list of prosite signatures which can be used to define and identify vertebrate homologues of UNC-53.

30 Figure 29 is a annotated sequence map of plasmid pEGFPsac. The GFP-C.elegans unc53sac fusion protein, and the C.elegans unc53 sac fragment are indicated.

 Figure 30 is a sequence map of plasmid pEGFP72. The GFP-C.elegans unc53 fusion protein and the C.elegans unc53 fragment are indicated.

35 Figure 31 is an annotated sequence map of plasmid

- 34 -

pEGFPsma. The GFP-C.elegans unc53sma fusion protein, and the C.e.unc53 sma fragment are indicated.

Figure 32 is an annotated sequence map of plasmid pEGFPec1. The GFP-C.elegans unc53ec1 fusion protein, and the C.elegans unc53 ec1 fragment are indicated.

Figure 33 is an annotated sequence map of plasmid pEGFPxba. The GFP-C.elegans unc53xba fusion protein, and the C.elegans unc53 xba fragment are indicated.

Figure 34 is an annotated sequence map of plasmid pLM4. Open reading frames of the hui-unc53/1 and GFP are indicated.

Figure 35 is a sequence map of plasmid pNP8.

Figure 36 is an illustration of microtubule association of C.elegans Unc53, shown in HepG2 cells, transiently transfected with pTB72, expressing C.elegans Unc53. panel A: microtubule staining of HepG2 cells, using YL1/2 panel B: C.elegans Unc53 staining, using rab4.

Figure 37 is an illustration of microtubule plus-end association in human cell lines transiently transfected with pTB72, expressing C.e.Unc53. C.elegans Unc53 was stained with mab-16-48. Panel C: COS cells showing microtubule association panel B: MCF7 cells showing microtubule plus-end association panel A: HepG2 cells showing microtubule plus-end association.

Figure 38 is an illustration of microtubule association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc53 fusion protein. GFP fluorescence was observed in living cells. Panel A: microtubule association of the GFP-C.elegans unc53 fusion protein panel B: microtubule plus-end association of the GFP-C.elegans unc53 fusion protein.

Figure 39 is an illustration of microtubule

- 35 -

association in N4 cells transiently transfected with pEGFP72, expressing the GFP-C.elegans Unc-53 fusion protein. Microtubules were stained with YL1/2 after paraformaldehyde fixation. Panel A: Microtubule association of the GFP-C.elegans unc53 fusion protein. Panel B: tubuline staining. Panel C: panel A plus panel B: co-localisation of the GFP-C.elegans unc-53 fusion protein and Tubuline can be seen as yellow.

Figure 40 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPsma, expressing the GFP-C.elegans unc53sma fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53sma fusion product. Panel B: Centriole association of GFP-C.elegans unc53sma fusion product when expressed at low levels.

Figure 41 is an illustration of microtubule association in N4 cells, transiently transfected with pEGFPec1, expressing the GFP-C.elegans unc53ec1 fusion protein. Panel A: Microtubule association of the GFP-C.elegans unc53ec1 fusion product. Panel B: Centriole association of GFP-C.elegans unc53ec1 fusion product when expressed at low levels.

Figure 42(a)/Figure 42(b) are illustrations of fluorescence of GFP in N4 cells transiently transfected with pEFPxba and pEFGPsac respectively.

Figure 43 is an illustration of microtubule association of in N4 cells transiently transfected with pLM4 expressing GFP-Hu-UNC53/1 fusion protein. Panel A: microtubule association of GFP-HU-UNC53/1 fusion protein. Panel B: microtubule plus-end association of GFP-Hu-UNC53/1 fusion protein. Panel C: microtubule association of GFP-Hu-UNC53/1 in dividing cells (end of division).

Figure 44 is an illustration of the sequence of Plasmid pNP9.

- 36 -

Figure 45 is an illustration of immuno
fluorescence in melanoma G361 cells stained with sera
28.1. Panel A: Microtubule plus-end association of
Hu-UNC53/1. Panel B: microtubule plus-end association
of hu1-Unc53 in growth cone extensions.

Figure 46 is an illustration of GFP fluorescence
and immunofluorescence in N4 cells transiently
transfected with pLM4, and stained with sera 28.1.
Panel A: Fluorescence of GFP-Hu-UNC53/1 fusion
protein. Panel B: Immunofluorescence of serum 28.1.

Figure 47 is an overview of the microtubule (+)
end, the microtubule and f-actin cytoskeleton binding
properties of different constructs

Figure 50 is an illustration of rescue of lateral
ALN neurons in mutant *unc-53*.

Dorsal view of the ALN neurones axones visualise
in GFP fluorescence with the transgene *pA/GFP* in the
posterior of an adult, (c) cellular body.

a) wild type, anterior axon (aa) migrates in a
straight line along the body until reaching the head,
on the dorsal sublateral cord, posterior axon (ap)
migrates into the tail;

b) *unc-53(n152)*, anterior axons are the shorter, stop
ahead of the vulva region and form numerous collateral
branches towards the dorsal cord;

c) *unc-53(n152), pA/unc-53* anterior axons no longer
form branches and migrate in a straight line into the
head, as in the wild type at a).

scale bar 10 μ m.

Figure 51a : is an illustration of chimeric
fusion between *C. elegans* and human 1 homologue of the
unc-53 gene. The region of the putative nucleotide
binding domain (NTP) is replaced in the *C. elegans*
cDNA by the same region of the human homologue 1 of
unc-53 (H1). The cDNA is under the promotor region A

- 37 -

(pA) of unc-53, which raise expression in the ALN lateral neurons.

Figure 51b : is an illustration of the chimeric minigene nematode/human pA/unc-53-H1 partially rescue the defect in the longitudinal migration of the lateral neurons ALN and PLN. The four strains compared are : wt; unc-53(n152); unc53(n152),pA/unc-53; unc-53(n152),pA/unc-53-H1. The observed phenotypes are put in three classes :

$\frac{1}{2}$ sauvage η , the axon is straight, unbranched, and migrates until the head; $\frac{1}{2}$ vulve η , the axon is straight, unbranched, and stops in the vulva region; $\frac{1}{2}$ mutant η , the axon is short, never joints the vulva region and made a lot of collateral branches. Numbers are in percentage. The number of observed axons are noted in the last column. The chimeric fusion between the C. elegans gene and human homolog (unc-53-H1) partially rescues the mutant phenotype. The chimeric gene was maded by replacing the putative nucleotide binding region (NTP)of the nematode cDNA by the same region of the human homolog 1 (H1).

Figure 52 is an illustration of the sequence for plasmid pLM5.

Figure 53 is an illustration of the sequence for plasmid pLM6.

Figure 54 is an illustration of the sequence for plasmid pLM1.

Figure 55 is a sequence map of plasmid pCB251.

Figure 56 is a sequence map of plasmid pNP10.

Figure 57 is a sequence map of plasmid pCB501.

Figure 58 is a sequence map of plasmid pTB115.

Figure 59 is a sequence map of plasmid pPD95.75.

Figure 60 is a sequence map of clone X16.

Figure 61 is a sequence map of plasmid pLM3

DEPOSITED MATERIALS

	Deposit	Date	Acc. Nr
	pCB201 plasmid DNA in E. coli	3 December 1996	LMBP 3594
5	Lambda clone 3B encoding hu-unc-53/1	3 December 1996	LMBP 3595
	MCF-7 clone z4 (mock)	3 December 1996	LMBP 1600CB
	MCF-7 clone (pCB201)	3 December 1996	LMBP 1601CB
	NIH-3T3 mock	3 December 1996	LMBP 1602CB
	NIH-3T3 pCB201	3 December 1996	LMBP 1603CB
10	pLM1	13 November 1997	LMBP 3762
	pLM4	13 November 1997	LMBP 3763
	pEGFP72	13 November 1997	LMBP 3764
	pCB501	13 November 1997	LMBP 3765
	BAC clone comprising fragment of hu-unc53/2 gene	15 November 1997	LMBP 3773
15	Worm strain with chimeric C.elegans/human unc-53 gene	15 November 1997	LMBP-1663CB

20

The above plasmids and cell lines were deposited at the Belgian Coordinated Collections of Microorganisms (BCCM) at laboratorium voor moleculaire biologie - plasmidencollective (LMBP) B9000, GENT, Belgium, in accordance with the provisions of the Budapest Treaty of 28 April 1977.

25

The present invention will now be described with reference to the following examples which are not limiting.

30

Identification of a human homologue of the UNC-53 protein of C.elegans.

Extensive searches with the ce-UNC-53 sequence (Figures 1 and 2) against the public domain databases

- 39 -

(EST, Genbank, EMBL, Swissprot and PIR) revealed no statistically significant homologies (a smallest sum probability (ssp) of $10 e - 8$ is generally accepted to be significant at amino acid level). Two ESTs
5 gbH09036 (ssp = $1.1 e - 5$) a Homo sapiens cDNA clone and gbAA049124 (ssp= $8.6-5$) a mouse cDNA clone showed homology to a "coiled coil" region a common motif in the contributing to protein secondary structure.
(figure 3)

10 All other candidate scores were are at background level (ssp >0.21). Careful examination of weak candidate ESTs identified EST gb:R41071 from Homo sapiens, which had obtained a low score of 53 and a non-significant probability score of 0.33 (Fig. 4).
15 The inventors surprisingly discovered potentially significant homology with the Ce-UNC-53 nucleotide binding domain, provided multiple frameshifts and sequence errors were hypothesized.

The inventors amplified, cloned and sequenced
20 part of gb:R41071 from human heart and human lung cDNA and from human genomic DNA and discovered that clone gb:R41071 had up to ten 10 different mistakes in the region checked. 5 extra nucleotides were scattered along its sequence and two nucleotide substitutions
25 were identified, and gb:R41071 lacked three nucleotides present in our clone (Fig. 5). The novel sequence obtained was two nucleotides shorter and showed the two UNC-53-homologous regions in frame. The genomic fragment obtained is larger (700 bp total
30 length) than the corresponding cDNA clones indicating the presence of an interverting sequence of around 500 bp in nucleotide 162 of this fragment. The amplified cDNA fragment which was cloned to vector PCRII (Intvitrogen) and named pCR231 and was used as a probe
35 to screen cDNA libraries.

- 40 -

The conceptual translation of the clones we obtained by PCR were screened using blast and tblastn against all known protein and DNA sequences in the database. The only clone which came up with statistically significant similarity was Ce-UNC-53 (Fig.6). This human clone and Ce-UNC-53 thus form a unique homologous pair compared to the rest of the known sequences, indicating the statistical relevance and novelty of our discovery. We designate this human gene as hu-UNC-53/1. Human heart and a human colorectal adenocarcinoma cDNA libraries were probed with pCR231 probe to identify longer cDNA clones. The clones overlap giving a linear sequence of 3706 bp (Fig 8 and 26). This sequence shows an 959 amino acid open reading frame from the beginning of the clone. The absence of a 5' untranslated region suggests that the mRNA will extend 5'.

Sequence alignment searches of the public domain databases with the DNA sequence of hu-UNC-53/1 and its' conceptual translation identified a series of ESTs most of which correspond to the 5' UTR region. (Figures 7 and 8). Surprisingly, hu-UNC-53/1 identified also the cDNA clones gbH09036 and gbAA049124 homologous to the predicted coiled coil region in Ce-UNC-53 hu-UNC-53/1, and furthermore identified a third weakly homologous EST gbR21023. The inserts of gbH09036, gbAA049124 and gbR21023 were obtained from the Merck consortium and sequenced.

gbAA049124 is >95% identical to Hu-UNC-53/1 over 604 available amino acids (fig. 10) and is the mouse orthologue of Hu-UNC-53/1. The insert in gbH09036 is clearly homologous to hu-UNC-53/1 but derived from a different locus. We therefore name the gene identified by gbAA049124 Mu-UNC-53/1 and the gene identified by gbH09036 Hu-UNC-53/2. (Figure 11).

- 41 -

5 domains of high similarity mark the *unc-53* gene family

5 Ce-UNC-53 and the here-identified vertebrate homologues form a unique novel protein family, that is distant from the remainder of the proteins in the public domain. Alignment of the predicted open reading frames shows that Hu-UNC-53/1 and Hu-UNC-53/2
10 are equidistant from Ce-UNC-53. The highest homology is found in the carboxyterminal amino acids of Ce-UNC-53 region. The presence of a conserved GXXGKS/T box suggests a nucleotide binding function. However, this domain as a whole does not belong to a class of known
15 nucleotide binding proteins.

 The similarity amongst the presently known sequence of the UNC-53 family of proteins is highest in 5 blocks over most of the available sequence (959 amino-acids) and a firther block identified in Figure
20 12a. These blocks can be assigned signature sequences as displayed in figure 28 or can be assigned weight matrices based on the alignment between the different family members. By using truncated constructs of Ce-unc-53, the functional relevance of these domains has
25 been addressed.

Hu-UNC53/1 and Hu-UNC-53/2 are complex transcription units.

30 1. A cancer cell line RNA blots probed with HU-Unc53/1.

 A Northern blot of poly-A+RNA from several cancer cell lines (Melanoma G361, Lung Cancer A549, Colorectal Adenocarcinoma SW480, Burkitt Lymphoma
35 DRajii, Leukemia Molt4, Lymphoblastic Leukemia K562,

- 42 -

HeLa S3 and Promyelocytic Leukemia HL60) was probed using the whole insert of pHH3b. No or weak expression was detected in the Burkitt Lymphoma DRajii, the Leukemia Molt4 and the Promyelocytic Leukemia HL60 cell lines. Five different transcripts are detected in the remaining cancer cell lines: transcripts 1 and 2 are larger than 9.5kb, transcripts 3 and 4 are 6 to 7 kb and the fifth transcript is around 6 kb. Transcripts 1 and 2 are present in all expressing cell lines. Transcripts 3 and 4 are restricted to Melanoma G361, Lung Cancer A549 and Colorectal Adenocarcinoma SW480 and are the predominant transcripts in Melanoma G361 and Colorectal Adenocarcinoma SW480. Transcript 5 is restricted to Lymphoblastic Leukemia K562 and HeLa S3 and is predominant in HeLa S3.

2. Cancer cell lines RNA blots probed with HU-UNC-53/2.

A similar set of cancer cell line Northern blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-aaacaccagtgagtcc. HU-UNC-53/2 is expressed in Melanoma G361, Colorectal Adenocarcinoma SW480, Lymphoblastic Leukemia K562 and HeLa S3. No expression was detected in Lung Cancer A549, Burkitt Lymphoma DRajii, Leukemia Molt4 and promyelocytic leukemia HL60. Interestingly only 2 transcript sizes were detected of around 7 kb expressed in Lymphoblastic Leukemia K562 and HeLa S3 and a transcript of >9.5 kb in Melanoma G361 and Colorectal Adenocarcinoma SW480.

3. Normal Human tissue probed with HU-Unc53/1.

A Northern blot of poly-A+RNA from normal

human tissue was probed using the whole insert of phage HH3b. Expression levels are low in all tissues with the highest level in heart and placenta, several fold lower levels in brain and testis, even lower
5 levels in skeletal muscle, pancreas, thymus, colon, small intestine, ovary and prostate. Expression in peripheral blood leukocyte, lung, liver, kidney, spleen is barely detectable.

10 4. Normal Human tissue probed with Hu-unc53/2.

A similar set of blots were probed with a 652bp fragment of EST46037 amplified by using the primers 5'-aggagatgaagctgacagatatcc and 5'-
15 aaacaccagtgaagtcc. Expression levels are low in all tissues with the highest level in kidney, lower levels in heart, placenta, lung, skeletal muscle and pancreas. Expression is barely detectable in brain and liver.

20 The hu-UNC53/1 and hu-UNC-53/2 homologues are clearly highly regulated genes, showing a strong tissue specificity and, probably, additional mechanisms of regulation (ie differential splicing of different promoters). The different proteins derived
25 from RNA's identified by probe hh15 presumably share the carboxyterminal nucleotide binding domain. Ce-UNC-53 was shown to be a complex genetic locus and complex transcription unit. The different transcripts are thought to be a mechanism to assure the necessary
30 specificity and functional diversity of this signal transduction pathway, with respect to different signals and receptors, different tissues and different directions of migration. The occurrence of a new transcript or the observed changes in expression
35 levels in the cancer cell line blot suggests a role

- 44 -

for hu-UNC-53/1 and hu-UNC-53/2 in the establishment or maintenance of the transformed state of those cells.

5 **Phenotypic changes in cells transfected with the Nucleotide Binding Domain of Ce-UNC-53/1 and Hu-UNC-53/1**

10 Ectopic expression of full length Ce-UNC-53 in *C. elegans*, murine neuroblastoma cells or human MCF-7 breast-carcinoma cells, has been found to lead to increased filopodia outgrowth and increased motility (unpublished). The structure of Ce-UNC-53 protein is reminiscent of that of large kinases or dynamin where
15 a catalytic domain is positively or negatively regulated by domains that interface with signal transduction pathways for example (by GRB2 binding, phosphorylation or the like). The inventors therefore decided to test whether the nucleotide domain by
20 itself is capable of inducing the observed changes in the microfilament cytoskeleton and motile or ruffling behaviour.

 cDNA fragments coding for the nucleotide binding domains of Ce-UNC-53 and Hu-UNC-53/1 were cloned in
25 mammalian expression vectors with the CMV promoter (see experimental procedures).

 To be able to detect expression from pCB201 (Fig. 13), an N-terminal his and a T7 epitope tag were fused in frame with the hu-UNC-53/1 cDNA hh15. pCDU3
30 contains a larger fragment of Ce-UNC-53 and starts just before the conserved "VIELKIEL" domain (Fig. 12).

 The empty pCDNA3 vector or pCDNA3.1-His-LacZ, a mammalian expression vector for *E. coli* Beta-galactosidase, was used as a control vector (mock
35 transfection). The differences between mock and

- 45 -

transfected N4 and MCF-7 clones were analysed using phase-contrast and Nomarski microscopy coupled with time lapse analysis, phagokinesis and immunocytochemical characterisation of the F-actin.

5

Phenotypic changes in mouse N4 neuroblastoma cells

10 N4 neuroblastoma cells were stably transfected with control construct pCDNA3.1 and the C. elegans UNC-53 constructs pTB72, pCDU2, pCDU3 and pCDU4. The population of clones transfected with the empty expression vector were homogeneous and similar to wild type N4 cells. In contrast thereto, 1/4 to 50% of the clones transfected with pTB72, pCDU2, pCDU3 and pCDU4 (see experimental procedures and Figs. 1,17,15 and 16 respectively) had distinct phenotypes:

1. Wild type or N4 cells transfected with pCDNA3, designated as mock transfection show a central cell body, with extensions, designated as neurite outgrowths. Less than 5% of the population have lamellae. When present, they are generally situated on the cell body and on the opposite site of the neurite extensions (figure 22a). The lamellae show a radial actin spike pattern. Limited branching of the actin fibres is observed in wild type or pCDNA3 transfected N4 cells. Side branches are smaller and can be clearly distinguished from the main actin branch (figure 23a).

2. N4 cells, stably transfected with pCDU4, harbouring the homology block E, show an overall morphology which is similar to that of wild type N4's (a cell body with neurite outgrowth). They exhibit however an increased frequency and level of lamellae formation (figure 22b). These lamellae, which contain

- 46 -

F-actin microspikes are found on both the cell body and the neurite outgrowth (figure 23b). Wild type N4 cells, in contrast thereto, rarely exhibit lamellae on the neurite outgrowths.

5 3. N4 cells, stably tranfected with pCDU3, encoding for homology blocks C, D and E, show an even higher level of lamellae formation labelled with TRITC-phalloidin, the cells appear surrounded with F-actin fibres, consisting of bundles of F-actin
10 microspikes (figure 23c). The presence of these lamellae has completely modified the general appearance of the cells. They appear flatter and in 90% of the population, it is not possible to distinguish between the cell body and the wide neurite
15 as they flow gradually into one another (figure 22c). If wild-type-like thin neurite-like outgrowths are present, they are frequently numerous, branched and located all around the cell.

20 4. The overall morphology of N4 cells, stably transfected with pCDU2, encoding for homology blocks A, B, C, D, and E, resembles that of the wild type cells since, cell body and neurite outgrowth can be clearly distinguished. The pCDU2 transfected cells however show more neurite outgrowth, and these are
25 long and very branched, especially at the end of the outgrowth. When neurite outgrowths of different cells make contact, increased branching can be observed, giving the appearance of a network (figure 22d). N4 cells, transfected with pCDU2, show bundles of long
30 radial F-actin filaments (microspikes), which can be branched, especially apically. The space between the hand-shaped actin spikes is mostly filled in with actin, leading to small lamellae-like structures. Also the network-like branching between the cells
35 shows both the bundled actin structures and the

- 47 -

lamellae-like fill-in features. These dense F-actin structures are sometimes seen on the cell body, which enhances the network-like appearance of the cells (figure 23d).

5 5. N4 cells, stably transfected with plasmid pTB72, encoding the full length C. elegans UNC53 protein, seem to have a more rigid structure than wild type cells, most clearly seen as spindle-like and triangle-like cells. The corners of these cells show
10 an increased level of hand-like lamellae structures. This specific phenotype is best seen when the cells are grown at low density (figure 22e, Fig. 23e).

15 Phenotypic changes in human breast carcinoma MCF-7 cells

MCF-7 cells were stably transfected with the pTB72, pCDU2, pCDU3, pCDU4 and pCB201. The population
20 of clones transfected with the LacZ-expression vector were homogeneous and similar to wild type MCF-7 cells. In contrast thereto, ~30-50% of the clones transfected with pTB72, pCDU2, pCDU3, pCDU4 and pCB201 had distinct phenotypes which were analysed as above for
25 the N4 cells:

1. Wild type and mock (pcDNA3) transfected MCF-7 cells are heteromorph. In general they are round cells or clusters of cells surrounded by lamellae. Bulges, similar to thick filopodia, can be observed
30 (figure 19a). When the cells are stained with FITC- or TRITC coupled phalloidin, F-actin actin stress fibres can be observed, often in rings surrounding the cell body (figure 20a & 21a). When cells are round up like this actin is present at the edge of the cell
35 body. Less than 10% of the cells display filopodia

- 48 -

filled with radial F-actin microspikes. In time-lapse analysis the cells are highly quiescent with limited ruffling at the edge of the cell.

2. MCF-7 cells transfected with pCDU4, encoding
5 for homology block E, show two major phenotypic
differences compared to the wild type cells. These
cells are more flat and have more extended
lamellipodia leading to a pancake-like appearance.
Some clones show more filopodia than wild type (figure
10 19b). Radially organised F-actin fibres can clearly
be observed in the lamellae surrounding the cells.
These stress fibres resemble the wild-type structures,
but have a more radial than circular orientation. In
the filopodia, one can observe an increase of
15 apparently unorganised, bundles of actin patches
(figure 21b).

3. MCF-7 cells, stably transfected with pCDU3,
encoding the homology blocks C, D, and E, shows a
strikingly different and constant morphology. The
20 cells appear smaller than wild type because they are
more rounded up. All the cells have more filopodia,
surrounding the cell body (figure 19c).
Morphologically these filopodia have the same "hand-
like" appearance as those observed in N4 neuroblastoma
25 cells. Such filopodia are hardly ever observed in
mock transfected MCF-7 cells. These filopodia are
filled with F-actin fibres. Compared to wild type
cells, fine actin stress fibres are decreased (figure
21c). In time-lapse analysis single cells as well as
30 clusters of cells can be seen to ruffle much more
dynamically than single or clusters of wild type
cells. The "half-life" of a filopodia outgrowth on
the cell surface is much shorter in transfected cells
and the numbers of filopodia present at any time
35 higher.

- 49 -

4. Cells transfected with pCB201 (which is structurally similar to pCDU4 but human) has a phenotype that is nearly indistinguishable from that of cells transfected with pCDU3 except that the observed phenotype and ruffling activity and filopodia outgrowth is even higher than pCDU3 (figure 18).

5. The overall morphology of the MCF-7 cells transfected with pCDU2, which encodes the homology blocks A, B, C, D and E, resembles that of the pCDU3 transfected cells. The cells are more rounded up and show more filopodia than the wild type and mock transfected cells (figure 19d). The filopodia, which are all around the cell body tend to be longer, and show a difference in actin organisation. The small filopodia have the same actin bundles as seen in the pCDU3 transfected cells. In the longer filopodia, the actin bundles are more parallel, and radial to the cell body (figure 21d).

6. MCF-7 cells transfected stably with pTB72, encoding the full length UNC53 protein, are extremely rounded up, and tend to adhere more than wild type cells. The cells grow in clusters with sausage- or tube-like shapes. The presence of large extremely thin lamellae with a surface area of more than three times the central cell body forms a second morphological feature, unique for the pTB72 transfected MCF-7 cells (figure 19e). These sheets are difficult to observe under a phase contrast microscope, but are very clear when stained with phalloidin. The lamellae protrude from one side of a cell or group of cells and are filled with thin long criss-crossing actin fibres, different from "giant" wild type MCF-7 cells (figure 21e).

These experiments lead to the following set of conclusions: (Figure 47 summarises the data of the

- 50 -

domain swapping experiments in *C. elegans unc-53*)

1. Murine and human cells transfected with the Ce-UNC-53 or hu-UNC-53/1 domains show clear effects on the nature and dynamics of their motile behaviour as demonstrated by changes in the F-actin cytoskeleton (the increase in lamellipodia, hand-like filopodia and "hair-like" microspikes on the cell surface and the associated reduction of the "rings of F-actin" stress-fibres).

2. This effect is found in two cell types of different species and tissue origin: MCF-7 cells (human breast carcinoma cells of epithelial origin) and murine N4 neuroblastoma cells. pCB201, pCDU3 and pCDU4 induce in MCF-7 cells a type of filopodium which is frequent in wild type N4 cells but rare to absent in wild type MCF-7 cells, suggesting the activation by these constructs of motile behaviour which is "normal" in N4 cells but of an unusual type in MCF-7 cells. This indicates the activation of a specific downstream process as opposed to a disruption of an existing process. It is well known that some cell types prefer to migrate with filopodia and other cell types with lamellipodia.

3. Expression of pCB201, pCDU3 and pCDU4 gives qualitatively similar F-actin remodelling and increased filopodia and lamellipodia outgrowth. pCB201 and pCDU3 are however much more active in this process than pCDU4.

4. pCB201 is a much more potent activator of filopodia outgrowth than pCDU4, which is to be expected considering the large evolutionary distance between *C. elegans* and vertebrates.

5. These experiments identify homology domain E (predicted nucleotide binding domain) of UNC-53 as the

- 51 -

"domain" that activates F-actin remodelling and filopodia/lamellipodia outgrowth. Progressive addition of the aminoterminal homology A,B,C,D lead to qualitative and quantitative modulation of the phenotype present in domain E.

6. Homology domains C and D (pCDU3) "enhance the basic activity present in homology domain E (pCDU4/pCB201).

7. Homology domains B and C (pCDU2) qualitatively modify the phenotype of domain E, leading morphologically different lamellipodia formation than pCDU3 transfected cells. It is thought that lamellipodia and filopodia formation are mediated by different signal transduction pathways requiring two related but different Ras-like G-proteins RAC for lamellipodia formation and CDC42 for filopodia formation.

8. pTB72 which includes homology domains A,B,C,D,E plus an additional 700 amino acids not yet identified isolated in the human members of the family confers a more localised filopodia outgrowth and a different morphology.

9. The expression levels of pTB72 (full length C. elegans UNC-53), pCDU3, pCDU4 and pCB201 are extremely low. The observed effect is therefore unlikely to be due to dominant negative effects (such as stoichiometric depletion of other cellular components) or structural changes in the actin cytoskeleton mediated by UNC-53 or its fragments.

The data point to a multi-domain organisation in UNC-53 whereby the aminoterminal domains exert positive (e.g. pCDU3) and negative (e.g. pCDU2) control on the activity of the domain E or are leading to novel activities or the localiation of the activity in the cell (pCDU2, pTB72). Our observation that the

- 52 -

nucleotide binding domains (NTB) of distantly related members of the UNC-53 family induce similar phenotypes, suggests a general role for this domain of the UNC-53 family.

5

CELLULAR ASSAYS TO IDENTIFY PHARMACOLOGICAL
MODULATORS OF UNC-53 AND COMPONENTS OF THE UNC-53
PATHWAY

10

Mammalian and human cells transfected with plasmid constructs containing unc-53 sequence of either C. elegans or of human origin were observed to display obvious, specific and similar changes in comparison to mock or untransfected parent cells. These changes relate to the functioning of the cytoskeleton, in particular the F-actin cytoskeleton, to cell locomotion and directionally cell motility and reflect UNC-53 gene family members as capable of playing an integrator function in cell motility.

15

20

The cellular tools derived through transfection and derived functional assays with these cells not only enable characterisation of the motile phenotype typically observed after introduction of unc-53 genes, they also can be easily adapted to screen for pharmacological compounds that interfere with either (1) the expression of unc-53 gene family members, (2) the cellular functioning of unc-53 transgene(s) and of components in the unc-53 signal transduction pathway.

25

30

Two classes of pharmacological modulators are envisaged.

35

A first class are inhibitors of UNC-53s or the unc-53 pathway(s), which revert the described phenotypic changes induced by unc-53 transgenes or

- 53 -

aspects thereof. Such compounds are considered relevant leads to target diseases where unwanted directional motility of cells occurs such as metastasis, angiogenesis or inflammation.

5 Secondly, pharmacological stimulators are envisaged, such as compounds which induce - in non-transfected cells - phenotypes that induce or mimic (aspects of) the described 'unc-53' phenotype. Such compounds may do so by inducing or upregulating
10 expression levels of a known unc-53 gene or by activating endogenous (yet unidentified) members of the unc-53 gene family. The target application here are wound and tissue repair, in particular diseases such as neuronal regeneration and plasticity.

15 The nature of compounds envisaged can be small (organic) molecules, bio-molecules (such as peptides, sense or antisense (oligo-)nucleotides or chemical modifications thereof. Alternatively, compounds can be thought of as a series of plasmid nucleotide
20 constructs containing gene sequences in a screen for novel unc-53-unrelated genes with a similar functional effect in the cell or genes related to the unc-53 gene family or novel members of the unc-53 gene family based on sequence similarity such as for example the
25 genes in plasmids pTB72, pcDU3, pcDU4, pcDU2, pcB201, or modifications thereof such as for example epitope tagged, deletion, complementation or mutagenised nucleotide constructs.

30 The cellular assays envisaged in the claims have been exemplified for three cell lines: the human breast carcinoma cell line MCF-7, the mouse neuronal cell line N4 and the mouse fibroblast cell line NIH-3T3. Pharmacological assays are focused on quantification of endpoints in a high throughout
35 screening mode. Many of the computer aids for

- 54 -

(semi-) automation are well known to the field and currently applied in the applicants labs. Given the subtlety of the phenotypes observed, primary focus was given to morphological assays that assess the phenotypes or aspects thereof.

The nucleotide binding domain of Hu-UNC-53/1 has transforming activity in NIH3T3 fibroblasts

Biochemical and genetic analysis suggest that UNC-53 functions in GRB-2 mediated signal transduction pathways controlling cell motility. The occurrence of an altered hu-UNC53/1 mRNA pattern in cancer cell lines, moved us to investigate if whether hu-UNC53/1 plays a role in the transformed state of those cells.

Thereto, we tested the ability of the nucleotide binding domain of hu-UNC-53/1 and Ce-UNC-53 to transform NIH/3T3 cells. Construct pCB201 (hu-UNC-53), which induces ruffling behaviour and cell motility, were transfected into NIH3T3 cells. Positive controls included Myc and H-ras. Negative controls included empty vector and Rac 1N17 and cdc42N17.

The cells that survived G418 selection were assayed for loss of contact inhibition (their ability to grow as foci). Positive controls included the combination of two well known oncogenes Myc and H-ras which were able to produce a high number of foci. The nucleotide binding domains of both Ce-UNC-53 and hu-UNC-53/1 are able to induce foci in this assay (Fig 24 & Table 1).

35

Table 1 Foci formation in NIH-3T3 cells stably transfected with pcB201

	mock	pcB201
5	22	138
	59	143

10

This suggests that the function of UNC-53 is not restricted to the activation of motility. UNC-53 may exert this additional function through the activation of as yet to be identified signal transduction pathways. Oncogenes frequently arise when a "controlling" domain and "activation" domain are separated through chromosomal rearrangements or integration of a part of a gene in the oncogenic virus. E.g. Erb Receptor tyrosine kinases, Ost a nucleotide exchange factor for Rac-1.

20

Hu-UNC-53/1 is localized to chromosome 1q31.1

Clone F226 (BACH-135 (014), Genome Systems, inc) was isolated from a human genomic BAC library using pCR231 as a probe and was confirmed by sequence analysis to be derived from the hu-UNC-53/1 locus. Purified DNA from clone F226 was labeled with digoxigenin dUTP by nick translation. Labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes derived from PHA stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextran sulfate and 2X SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluoresceinated antidigoxigenin antibodies followed by counterstaining with DAPI. The initial experiment resulted in

25

30

35

- 56 -

specific labeling of the long arm of a group A chromosome. A second experiment was conducted in which an anonymous probe which was previously mapped to 1p34 and confirmed by cohybridization with a chromosome 1 centromere specific probe, was cohybridized with F226. The experiment resulted in the specific labeling of the long and short arms of chromosome 1. Measures of 10 specifically hybridized chromosomes 1 demonstrated that F226 is located at a position which is 52% of the distance from the heterochromatic-euchromatic boundary to the telomere of chromosome arm 1q, and that corresponds to band 1q31. At total of 80 metaphase cells were analyzed with 72 exhibiting specific labeling (Fig. 25).

Gains of DNA sequences in 1Q31 were found in more than 10% of primary bladder tumors (Genes Chromosom Cancer 12: 213-219 (1991)). A putative tumor suppressor gene located near the locus F13B on chromosome arm 1q31-q32 appears to be involved in the pathogenesis of medulloblastoma (Int. J. Cancer 67: 11-15 (1996)). Loss of heterozygosity in this region of chromosome I has been implicated in development of human hepatoblastoma. Partial trisomies of 1q31 were found in Ewing's Sarcoma cell lines isolated from patients Cancer Genet Cytogenet 12: 1-19 (1984).

HU-UNC-53/2 is localised to Chromosome 11p15.1

DNA from clone F329 from BAC for Hu-unc-53/2 was labeled with digoxigenin dUTP by nick translation and applied in the experimental settings used for FISH of Hu-unc53/1 with F226. The initial experiment with F329 resulted in the specific labeling of the mid short arm of a group C chromosome which was believed to be chromosome 11 on the base of size, morphology

- 57 -

and banding pattern. A second experiment was conducted in which a biotin labeled probe specific for the centromere of chromosome 11 (D11Z1) was cohybridised with clone F329. This experiment
5 resulted in the specific labeling of the centromere in red and the mid short arm in green of chromosome 11. Measurements of 10 specifically labeled chromosomes 11 demonstrated that F329 is located at a position which is 65% of the distance from the centromere to the
10 telomere of the chromosome 11p, an area which corresponds to band 11p15.1. A total of 80 metaphase cells were analysed with 72 exhibiting specific labeling.

15 Chromosome 11p15 is a region showing loss of heterozygosity (LOH) in a variety of human malignancies, primarily breast cancer (Ali et al., Science 238, 185-188 (1987); Winqvist et al., Cancer Res. 53, 4486-4488 (1993)) but also Wilms' tumor
20 (Dowdy et al., Science 254, 293-295 (1991); Cowell et al., Br.J.Cancer 67, 1259-1261 (1993)), ovarian and testicular malignancies (Lothe et al., Genes Chromosomes Cancer 7, 96-101 (1993); Weitzel et al., Gynecol Oncol. 55, 245-252 (1994)) stomach cancer
25 (Baffa et al., Cancer Res. 56, 268-272 (1996)), lung cancer (Ludwig et al., Int.J.Cancer 49, 661-665 (1991); Fong et al., Genes Chromosomes Cancer (1994)), infantile tumors of adrenal and liver (Byrne et al., Genes Chromosomes Cancer 8, 104-111 (1993)). Since
30 LOH is believed to indicate inactivation of a tumor suppressor gene at the location where LOH occurs, the frequent LOH found at 11p15 in multiple human cancers suggests the presence of either a cluster of tumor suppressor genes or a single tumor suppressor in this
35 region (Seizinger et al., Cytogenet. Cell genet. 58,

- 58 -

10080-10096 (1991)). Chromosome transfer studies have shown that chromosome 11 can suppress tumorigenicity of both human breast cancer (Negrini et al., Cancer Res.55, 3003-3007 (1995)) and Wilms' tumor cells (Dowdy et al., Science 254, 293-295 (1991)) and a gene (named HTS1 or ST5) that may be responsible for suppressing tumorigenicity in HeLa cells has been mapped to 11p15 (Lichy et al., Cell Growth Diff. 3, 541--548 (1992)). Abnormalities at 11p15 have also been identified in a variety of other cancers, including lung cancer (parental origin of 11p15 deletion) (Kondo et al., Oncogene 9, 3063-3065 (1994)), bladder cancer (Presti et al., Cancer Res. 51, 5405-5409 (1991)), myeloid leukemia (translocation) (Nakamura et al., Nat. Genet. 12, 154-158 (1996)), malignant astrocytomas and other primitive neuroectodermal tumors (deletions) (Fults et al., Genomics 14, 799-801 (1992)), rhabdomyosarcoma (Scrabble et al., Nature 329, 645-647 (1987)) and hepatocellular carcinoma (Fujimori et al., Cancer Res. 51, 89-93 (1991); Wang et al., Cell Genet. 48, 72-78 (1988)). Recently a gene, TSG101, was cloned that is mutated in human breast cancer and deleted in uncultured primary human breast carcinomas (Li et al., Cell 88, 143-154 (1997)).

DIAGNOSTIC ASSAY USING THE DNA SEQUENCE OF HUMAN UNC-53S

The differential expression of human unc-53 transcripts in Northern blots of normal tissues versus transformed cell lines and the chromosomal locus of hu-unc-53/1 at 1q31 being a locus linked to three diseases, suggests the potential implication of hu-unc-53 genes in oncogenesis. By using the complete

- 59 -

DNA sequence of hu-unc-53/1 or /2 or fragments thereof in FISH, the potential involvement of these genes can be diagnosed in patients as exemplified in figure 26. Alike, the use of these hu-unc-53 sequences in diagnostic PCR assays can be used to determine overexpression of hu-unc-53s or fragments thereof.

Assay for microscopic phenotypic UNC-53 transfected MCF-7 cells

Mock and unc-53 transfected MCF-7 cells were seeded at low density in culture plates and allowed to adhere to the vessel. Light microscopic inspection at different time points either on live cells or after chemical fixation with Karnovsky's fixative revealed that in pcB201, MCF-7 transfected cultures a rounded shaped cell body with at their boundaries many filopodia. In contrast, mock or untransfected clones had a predominant 'flat' phenotype - with little or no filopodia. Quantitative measurements confirmed the statistical significance of this shift in phenotype (table 2 below).

TABLE 2

Quantification of phenotypic changes in unc-53 transfected MCF-7 cells (*)

Transfection:	clone	no feet (**)	with feet (**)	fraction with feet
mock	e	34	8	0.19
		37	0	0
pcB201	2	17	92	0.84
		37	83	0.69
	16	27	62	0.70
		20	71	0.78
		13	85	0.87

(*) Clones were passaged thrice, frozen and stored.

- 60 -

Thawed cells were trypsinised at confluency, monodispersed, seeded in flasks and allowed to attach to substrate overnight to 48 hours. Cultures were fixed with Karnovsky fixative and inspected using phase contrast microscopy. In parallel experiments, resistance to genitacin was confirmed.

(**) values are expressed as cells per microscopic view.

10 Assay for ruffling and motile behaviour using automated time lapse

The dynamic changes in cells are well known in the field. Animations of e.g. actin ruffles in astrocytoma cells or of actin based cell motility in e.g. fibroblasts can be accessed (<http://www.stc.cmu.edu/CLMIBhp/Imggallpg/Moviespg/actinruffle.mov>) or (<http://util.ucsf.edu/mitchi/Movies/migration.html>) on the world wide web. The dynamic changes as a result of transfection with unc-53 can best be appreciated in time lapse video sequences. At high magnification, the 'filopodia' display arrays of microspikes with highly dynamic behaviour. A rough visual estimate suggests these phenomena to be at least 10-fold increased in pcB201 transfected cells relative to the mock-transfected MCF-7 cells. Animations of these clones in NIH-Image can be requested from author or applicant.

Time lapse video imaging probably is the most informative way to appreciate the unc-53-induced phenotype in MCF-7 and is amenable to high throughput screening in a pharmacological context. Time lapses compressing 5 minutes real time supply sufficient information to quantitate the intensity of the motile

- 61 -

behaviour of pcB201 transfected MCF-7 cells in e.g. 12 well plates. In addition, algorithms have been described in the field which can automatically compute the 'motile area' of cells by comparing cells in two
5 images appropriately spaced in time (van laerebeke et al., 1992, cytometry, 13, 1-8).

Assay for visualising unc-53-induced F-actin
recruitment in MCF-7 cells

10

Cultures were chemically fixed, detergent extracted and fluorescently stained for F-actin (filamentous-actin) using fluorescently labeled phalloidin (Wieland et al., 1985, Int. J. Peptide &
15 protein Res, 21, 3-10) which display in a more specific way the dramatic phenotypic changes to transfection with unc-53 transgenes. By using image capturing and analysis of the F-actin patterns, image analysis algorithms well known in the field can assess
20 in an automated way, the f-actin filament positions, texture and distribution relative to the nuclear position or gravity point of the cells. Such algorithms are capable of discriminating phenotypic changes and thus also effects of pharmacological
25 inhibitors of transgene-induced phenotypes as well as compound induced unc-53 like phenotypes in mock or untransfected cells.

Phagokinesis assay for unc-53-induced
30 directionality and quantity of motility

The methods are described in the experimental section. Two cell populations with different motile behaviour in phagokinesis assays were observed. In
35 table 3 below the fraction of mock and UNC-53

- 62 -

transfected MCF-7 cells that produced linear tracks in the phagokinesis assay are shown. In the mock transfected MCF-7 cells, 61% of the cells produce a round track (long and short axis less than 2-fold different) and 39% cells produced 'linear' tracks (long and short axis more than 2-fold different). pcB201 transfected MCF-7 cells produced an increase of the fraction of cells displaying 'linear' tracks to 50%. An increase in the fraction linear tracks was made for MCF-7 cells transfected with full sequence Ce-unc-53.

In addition, a significant increase of 50% in the median area of tracks of a culture vessel was observed in the pcB201 transfected MCF-7 cells versus mock transfected MCF-7 cells (Table 2). These observations suggest that pcB201 as well as pTB72 transfection into MCF-7 cells is capable of increasing *in situ* locomotion in Ce-UNC-53 MCF-7, e.g. by increasing spreading, ruffling, or other forms of non-directional motility in the 'round' population. In addition the Ce-UNC-53 transgene in MCF-7 cells drives a fraction of the MCF-7 cells from non-directional motility (round tracks) into directional migration (linear tracks). Clone 2 thus provides a tool to analyse inhibitory or stimulating effects of pharmacological compounds on directionality or quantity of cell motility in relation to UNC-53.

30

35

Table 3. Analysis of motility in phagokinesis assays

Track morphology: fraction linear tracks

5	plasmid	clone	round	linear	l/r
	Mock	z4	18	13	0.42
10			17	11	0.39
			22	12	0.35
	pCB201	Clone 2	16	9	0.36
			13	13	0.5
15			7	8	0.53
			9	9	0.5
Track Size					
20	Clone	average \pm SD	min	max	(N)
	Z4	1626 \pm 188	1444	2011	(8)
	Clone 2	2326 \pm 283	1989	2816	(8)

25 Assays for the localisation of unc-53 in the cell
to microtubules or microtubule (+) plus ends

UNC-53s have been shown to reside on microtubules and preferentially on the microtubule (+)-ends of cells. This localisation represents an important feature of the UNC-53 family of proteins, which is rarely observed in other proteins. Absence of microtubule (+)-end binding in the protein APC following mutation has been implied in the role of APC in colon cancer (Smith et al., 1994, Cancer Res., 54, 3672). In analogy, it can be postulated that the proper functioning of UNC-53 also may depend on its specific localisation in the cell.

40 The methods used in the examples which prove the co-localisation with microtubules form a base for a series of assays for compounds which specifically

- 64 -

affect microtubule (+)-end binding of UNC-53s. To the skilled eye, the typical localisation of an UNC-53 protein on microtubules can be readily recognised and thus is sufficient for the interpretation of whether the treatment with a compound has affected the localisation of this UNC-53 (or a fragment thereof). Moreover, by combining the described methods (co-localisation) - well known to one skilled in the field and exemplified by the methods in the "experimental procedures" section - one can unequivocally confirm a compounds ability of abrogating (or promoting) microtubule and microtubule (+)-end binding.

Such an assay comprises contacting a cell culture of a cell line expressing an UNC-53 with a compound in the culture conditions proper for the said cell line, followed by an incubation and finally observation of the UNC-53 (or fragment) in situ by e.g. fluorescence microscopy (for GFP-chimeras) or by fixing the cell culture and performing an immunocytochemical staining for the UNC-53 (or fragment). For the co-localisation, methods such as immunocytochemistry for the microtubules of a cell or cell line combined with either immunocytochemistry for Ce-UNC-53 or Hu-UNC-53s or fluorescent detection GFP-UNC-53 chimeras are performed consecutively.

C.elegans-UNC-53 preferentially binds microtubule plus-ends or GTP-tubulin

Biochemical characterisation of UNC-53 has shown that UNC-53 binds the SH3 binding domains of SEM-5/GRB-2 and binds F-actin *in vitro*. GRB2 has been localised to the cortex of the cell and reported to be involved in the control of cell motility. To determine the *in vivo* subcellular localisation of Ce-

- 65 -

UNC-53, we transiently transfected COS, HepG2 and MCF7 cells with pTB72, an expression construct containing the full length Ce-unc-53 cDNA. This construct was previously shown to activate cell motility in N4 neuroblastoma and MCF7 cells. This construct gives high transient expression in COS cells, high to medium levels of expression in MCF7 cells and medium to low levels of expression in HepG2 cells. To visualise UNC-53, tubulin and F-actin, transfected cells were stained with various combinations of the anti-Ce-UNC-53 mab 16-48-2, rabbit anti-UNC-53 polyclonal, anti-tubulin mab YL1/2 and fluorescently labelled phalloidin.

At high levels of expression UNC-53 co-localises with the entire microtubule cytoskeleton, but at lower expression levels UNC-53 signal is restricted to the terminal regions of the microtubules at the plus ends. Very low levels of the expression yield a dot-like pattern in the vicinity of the cortex of the cell.

To map the MTB plus end domain of Ce-UNC53, we made two constructs pcDU2 (figure 17) and pcDU3 (figure 15) in which the aminotermus of Ce-UNC-53 is deleted. Proteins corresponding to these constructs are thought to be made *in vivo* from different unc-53 promoters. Transient transfections followed by immunolocalisation showed these proteins to be cytoplasmic. In stable transfections in N4 neuroblastoma cells and MCF7 cells they were shown to be no longer toxic to cells but cause highly increased activation of filopodia formation. We thus uncoupled (1) toxicity of Ce-UNC-53 from activation of motility and (2) microtubule binding from the activation of motility.

- 66 -

Analysis of the microtubule association of the
C.elegans and Human 1 UNC53

To isolate the microtubule association domain of
5 the C.elegans UNC53, N-terminal GFP fusions were made.
C-terminal deletions on the fusion product revealed
that the microtubule association was localised in the
N-terminal half of the protein. A GFP fusion was also
constructed with the Human1-UNC-53, to analyse the
10 microtubule association properties of this protein.
The association with microtubules was confirmed. A
mouse anti sera was used to show the presence of
native Unc-53 on microtubule plus ends of melanoma
line G361. The epitope recognition of the antibody
15 was confirmed by immunohistology experiments with
mammalian cells, transiently expressed with pLM4,
expression the GFP-hu1-UNC53 fusion protein.

Results

20 1. When transiently transfecting pTB72 in
several cell lines C.elegans UNC-53 associates with
microtubules and preferentially the plus-ends of the
tubuline fibres. Transfection of plasmids pCDU3 and
25 pCDU2 in N4 and MCF7 cell lines did not result in the
observation of microtubule co-localisation. pCDU4
resulted in no staining using mab 16-48 antibody (LMBP
Accession No. 1383CB) concluding that the epitope for
this antibody is localised outside the fragment
30 expressed by pCDU4.

It is possible that the microtubule associated
domain is situated in the N-terminus of the protein.
For this reason, we constructed an N-terminal GFP
fusion with the full length C.elegans UNC-53 sequence,
35 and various C-terminal deletion derivatives. These

- 67 -

fragments encode the N-terminal part of UNC-53 from 139 to 760 aa.

Furthermore, to analyse if the cloned fragment of hui-unc53 also could be associated with microtubules, a plasmid encoding a GFP fusion with the hui-Unc53 protein was constructed, and introduced into mammalian cells. A derivative of this construct was also constructed.

2.

a) Transient expression of C.elegans Unc-53 GFP fusion in N4 neuroblastoma lines

N4 cells were transiently transfected with pEGFP72, encoding a fusion protein of GFP and full length C.elegans unc-53 sequence. On an inverted microscope, the fluorescence of the GFP molecule could be followed in living cells. Cells which expressed low to medium levels of the fusion molecule showed a normal morphology after 18h to 30h. In these cells the co-localisation of the GFP fusion protein with the microtubules could clearly be demonstrated (figure 38a). In cells which demonstrated a low but still distinct GFP fluorescence, specific microtubule plus-end association could be observed (figure 38b). Cells expressing high levels of the GFP fusion protein tend to round up, in such a way that the microfilaments are difficult to visualise. After 48h, almost no GFP expressing cells can be found. It has previously been observed in transient expression of Unc-53, using plasmid pTB72, that the protein is toxic for the cells. The transient transfection experiments with the pEGFP72 plasmid gives the same observation, indicating that at least two features of the Unc53 protein are conserved in the GFP fusion protein, being

- 68 -

the microtubule association and the toxicity of the protein.

The transfected cells were fixed with paraformaldehyde, and the tubuline was stained using antibody YL1/2 and antimouse-CY3 (Jackson Labs). Although a significant loss of GFP fluorescence was observed, one could clearly demonstrate that the filaments observed with the GFP fluorescence co-localise with the microtubules staining (figure 39).

Putative Assay

Mammalian cells, in this case N4, were transfected with a lipofecting agent (lipofectAMINE) while in suspension, not being attached to a surface. After transfecting those cells with pEGFP72, the transfected cell suspension could be diluted in 24- and/or 96-well plates, enabling them to attach to the surface. Each well may contain a different compound of the collection to screen. After 24h, plates could be automatically screened for fluorescence levels. Wells containing a compound that abolish the toxicity of the GFP-C.elegans UNC-53 fusion protein will give high levels of fluorescence. Compounds having no effect on the fusion product will give no or only low levels of fluorescence.

b) Transient expression of the truncated GFP-C.elegans UNC-53 fusion proteins.

To assay if the microtubuline association did occur in the N-terminal part of the C.elegans Unc-53 protein, various C-terminal deletions were constructed.

Transfection of pEGFPsma and pEGFPec1 coding

- 69 -

for 760 AA and 670 of the N-terminal part of C.elegans UNC-53 in fusion with GFP, resulted in microtubuline association, as could be visualised in living cells. The association with the microtubules is less abundant than observed when expressing the full length C.elegans UNC-53 protein, but fibres could clearly be observed (figures 40a and 41a). More background fluorescence is seen. This could be due to a lesser association to the microtubules or to a instability of the fusion protein. The association with microtubules could not be observed after fixing the cells with paraformaldehyde nor with methanol fixation, giving an extra indication for the weak association with the microtubule network of these proteins or potential instability of the fusion protein. At low expression levels the association of the GFP fusion protein with the centrosomes could clearly be detected (Figures 40b and 41b). Centrosomes are the location in the cell with the highest microtubule concentration.

No plus-end associations could be observed with the deletion constructs, even when cells where expressing low levels of the GFP fusion proteins. In the case of very low expressions, the centrosomes could clearly be detected.

When transfecting N4 cells with pEGFPsac or pEFPXba, coding for 139 aa and 256 aa of the N-terminal part of C.elegans UNC-53 in fusion with GFP, no microtubule association could be observed. This indicates that at least 670 aa of the N-terminus of the C.elegans UNC-53 is needed to have microtubule association (figures 42a and 42b).

c) Transient expression of the GFP-hu-UNC-53/1

- 70 -

fusion proteins and a deletion derivative.

Plasmid pLM4 was transiently transfected into N4 neuroblastoma cells, and GFP fluorescence was observed in living cells. GFP fluorescence of the available sequence of hu1-UNC-53 in fusion with GFP was localised at the microtubule level. Moreover, at lower expression levels, both the centrosomes, and specific plus-end association could be observed. As has been observed with the C.elegans UNC-53 derivatives in fusion with GFP, expressed by the plasmids pEGFPsma and pEGFPec1, the GFP association seems to be less tight as was observed by the full length C.elegans UNC-53 fragment in fusion with GFP. The observed instability of the fusion protein can be due to a lesser association to microtubules, or to a degradation of the fusion protein (figure 43).

d) Immunofluorescence on melanoma line G361, and on neuroblastoma line N4 transiently transfected with pLM4.

Introduction

Northern experiments show that the melanoma cancer line G361 expressed abundantly both the Human1 and Human2 homologue of C.elegans UNC-53. To test if the proteins could be localised in this cell line, a collection of mouse sera was tested on this cell line. To verify if the observation was due to a hu-UNC-53 recognition, and not to an artifact, a positive sera was applied to N4 cells transiently transfected with pLM4, expressing the GFP-hu1-Unc fusion.

- 71 -

result

a serum, designated 28.1 from a mouse previously injected with peptide (DNRTLPPKKGLYRY) a conserved sequence of the UNC-53 family was used for a immunolocalisation experiment on G361 cells fixed with paraformaldehyde. Antimouse-cy3 was applied as second antibody. Association with microtubule plus-end could clearly be observed. Moreover, in cells showing directional movement, observed as growth cones extensions, abundant staining can be seen in the tip of the growth cone (figure 45). To test whether the recognition of the microtubule associated protein was identical to the Hui1-UNC-53 protein, N4 cells were transiently transfected with plasmid pLM4 and consequently fixed with paraformaldehyde and stained with serum 28.1. Only cells that were transfected showed staining with 28.1, indicating that the antibody of 28.1 recognised the Hui1-UNC-53-GFP fusion protein (figure 46). This confirms that the staining of the microtubule plus-ends in the growth cones of G361 by serum 28.1 is due to a recognition of at least the Human1 and/or the Human2 homologue. It is concluded that the overexpression of the human homologue of C.elegans UNC-53 in the melanoma cancerline G361 is located on the microtubule plus-ends.

Conclusions

30

- a) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows Unc53 activity
- b) - GFP-C.elegans UNC-53 fusion protein expressed by pEGFP72 shows microtubule association
- 35 c) - GFP-C.elegans UNC-53 fusion protein

- 72 -

expressed by pEGFP72 shows microtubule plus-end association

5 c) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPec1 show microtubule association.

d) - GFP-C.elegans-UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPsma and pEGFPec1 no not show microtubule plus-end association

10 e) - GFP-C.elegans UNC-53-(deletion variant) fusion proteins expressed by plasmids pEGFPxba and pEGFPsac no not show microtubule associations.

f) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule association.

15 g) - GFP-hu1-UNC-53 fusion protein expressed by plasmid pLM4 shows microtubule plus end association.

i) - serum 28.1 recognises the Hu1-UNC-53-GFP fusion protein as expressed by plasmid pLM4 in transiently transfected Neuroblastoma cells N4.

20 j) - the expressed human homologue of C.elegans.- UNC-53 in melanoma line (being at least hu1-Unc-53) is associated with the microtubule plus-ends.

EXPERIMENTAL PROCEDURES

25 Materials

The oligonucleotides used in the PCR-RACE experiments were synthesised by Eurogentee (Belgium). Radioactive compounds were obtained from Amersham.

30 The pCDNA3.1 eukaryotic expression vectors, human 1GT10 cDNA libraries, marathon-RACE cDNAs, human, Northern blots and the T7-tag monoclonal antibody were purchased from Invitrogen. N4, MCF7 and NIH 3T3 cells were retrieved from the Janssen Research cell bank.

35

- 73 -

PCR-RACE conditions

1. A quick screen human cDNA library panel was used to amplify EST clone gb..R41071. The primers used
5 were ESTfw 5'-AATGGCTTCCTGGTTACCTGAG-3' and ESTrv 5'-CAAGTCAGCACCCCGAAGCAGCTCT-3'. Human genomic DNA was used also as template (100ng/reaction). The amplification conditions were as follows: 1 min at 94°C, 30 sec at 55°C, 30 sec at 72°C, then 35 more
10 times and a final extension of 20 min at 72°C. This PCT fragment was cloned in vector pCR2.1. The resulting plasmid was designed pCR231.

A human heart clone was also produced by RACE-PCR from a human heart Marathon cDNA using the following
15 conditions; 1 min at 94°C, 30 sec at 70°C, 3 min 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72°C. KlenTaq DNA Polymerase was purchased from Invitrogen.

For the mouse homologue, total RNA was obtained
20 from N4 murine cells as described. A first strand cDNA was synthesized from 2 µgr of RNA using Ready To-Go cDNA kit (Pharmacia). The primers used were M-ESTfw 5'CCTCTGTGGGCACCGAGGTCACC--3'. The amplification conditions were as follows: 1 min at 94°C, 30 sec at
25 58°C, 30 sec at 72°C, then 35 more times and a final extension of 20 min at 72°C. All the amplifications product were subcloned in pCRII-1 and several independent clones were analyzed by sequence.

30 2. Screening of Human Heart/Colorectal Adenocarcinoma cDNA library

A human heart cDNA library and a human colorectal adenocarcinoma cDNA library were screened using
35 pCR231bp as probe by the standard plaque hybridization

- 74 -

method. The screening produced several positive clones in each library called respectively λ HH3, λ HH4, λ HH15, λ CAD14 and λ CAD27. The positive phages were purified by two additional rounds of plaque screening and were then amplified.

3. 5' extension using PCR

Three primers with homology to the 5' end of clone λ HH3b were made:

HU53rv1 (5'-cct-ggg-act-gaa-gct-ggt-acc-tga-gcc-3'), HU53rv2 (5'-ttg-gga-aga-gtg-ttc-cga-tcc-cgc-tg-3') and HU53rv3 (5'gtt-gcc-cag-ctc-tgg-ggc-ttc-cac-tcc-3') and used together with λ gt10rv primer (5'-gag-gtg-gct-tat-gag-tat-ttc-ttc-cag-ggt-a-3') in three nested PCR reactions on a cDNA amplified library from Human Heart (Clontech). The reaction mixes contained 25pmol of each primer, 1 mM of each dNTP, 1 μ l KlenTaq Polymerase Mix (50x) and 0.1 ng DNA. The cycling parameters for the first PCR were: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 51°C and 3 min at 72°C and a final extension of 10 min at 72°C, using HU53rv1 and λ gt10rv as primers. 0.4 μ l of this primary PCR product was amplified using HU53rv2 and λ gt10rv as nested primers with the following parameters: 3 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 52°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. The second nested PCR reaction was performed on 0.4 μ l of a 1/50 diluted purified 2.4 kb fragment using HU53rv3 and λ gt10rv as primers: 3 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 56°C and 3 min 30 sec at 72°C and a final extension of 10 min at 72°C. A 774 kb amplification product was subcloned in pCR2.1, resulting in plasmid pCB210-14. The clone fragment was analyzed by sequencing. This fragment

- 75 -

extends 699 bp in 5' direction (see fig 9).

4. 5' extension using PCR

5 Primer HU53rv4 (5'-ccc-tgc-ttg-gtg-ctg-agg-aga-
ctg-g-3') was designed on the 5' end of clone pCB210-
14 and was used together with λ gt10rv to amplify a
fragment of the Human Heart cDNA library with the
following parameters: 3 min at 94°C, 35 cycles of 1
10 min at 94°C, 1 min at 60°C and 3 min 30 sec at 72°C
and a final extension of 10 min at 72°C. A 887 bp
fragment was subcloned in pCR2.1, resulting in plasmid
pCB212. The clone fragment was analyzed by
sequencing. This fragment extends a further 767 bp in
15 5' direction (see fig 9).

5. Human Heart Library screening using the 0.8 kb insert of pCB212 as probe

20 The EcoRI digested and purified clone pCB212 was
used as probe to screen the Human Heart cDNA library
(Clontech) using standard plaque hybridization method.
The positive phages were purified by two additional
rounds of plaque screening. The insert of the λ DNA
25 (produced using Qiagen Lambda Kit) was analyzed by
sequencing. This pHH14-3 resulted in a 2663 bp
fragment overlapping pCB212, pCB210-14 and the 3' end
(434 bp) of λ HH3b and in a 761 bp 5' extension (see
fig 9).

30

3' and 5' extension of HU-Unc53/2 from EST46037

WashU-Merck EST 46037

35 Transformed cells carrying the EST 46037 sequence
were ordered from Research Genetics. Plasmid DNA was

- 76 -

isolated using standard protocols (Qiagen plasmid DNA isolation kit), the sequence of the insert was determined.

5 3' extension of EST 46037 by RACE

Marathon-Ready cDNAs (Clontech) are premade "libraries" of adaptor-ligated double-stranded cDNA ready for use as templates in RACE experiments.

10 Five ml Marathon-Ready cDNA was used as template in a regular 50ml RACE. The RACE mixture contained 1x KlenTaq PCR buffer, 0.2 mM of each dNTP, 1x advantage KlenTaq polymerase mix (Clontech), 0.15 mM AP1 adaptor primer and 0.15 mM RACE gene specific primer. The
15 amplification conditions were as follows :

94°C for 1 min, 5 cycles of 94°C for 30 s and 72°C for 4 min, 5 cycles of 94°C for 30s and 70°C for 4 min, 25 cycles of 94°C for 30 s and 68°C for 4 min.

20 One-hundred-fold diluted RACE product was used as a template in a nested PCR with AP2 adaptor and gene specific nested PCR primers. Specific nested PCR fragments were cloned into pCR Γ 2.1 (TA cloning kit, Invitrogen) and the sequences of the inserts were
25 determined.

gene specific primer (EST46037-F1)

5'AGTGAGAACAATGCTGTGGACATGC nested gene specific
primer (ES46037-F2) 5'CTGCTCAACTGCAAGTACCACAAATGC
Marathon cDNA library : human placenta

30

WashU-Merck EST 923793

Transformed cells carrying the EST 923793
sequence were ordered from Research Genetics. Plasmid
35 DNA was isolated using standard protocols (Qiagen

- 77 -

plasmid DNA isolation kit), the sequence of the insert was determined.

5 RACE fragments 1.4 and 3.7, 5' extension of
 EST46037

Method as described previously. Gene specific
primer (EST46037-R1) 5'ACTGCCTTGAGACTCTGACTTCAGC
nested gene specific primer (ES46037-R2)
10 5'TGGGCAGAACTGAGAGCTTCTAAGC Marathon cDNA library :
human placenta

RACE fragments B2.1, D2.1, H2.1; 5' extension

15 Method as described previously:gene specific
primer (97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG
nested gene specific primer (97010708)
5'ACCTGAGTCCTTTCTTAGGCAAAGTGTTCC Marathon cDNA library
: human placenta (fragment B2.1)

20 human HeLa S3 (fragment D2.1) human colorectal
adenocarcinoma SW480 (fragment H2.1)

PCR fragments E2.3, C2.3

25 EST 485068 is similar to but not identical with
the 5'end of HU-Unc53/1. A primer pair consisting of
one 3' EST 485068 primer and one 5' HU-Unc53/2 primer
were used to PCR amplify those fragments. 1gt10 human
30 placenta Quick screen library (fragment C2.3) or
Marathon cDNA from human HeLa S3 (fragment E2.3) were
used as templates in a PCR. A 50 ml reaction mix
contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM
MgCl₂, 0.2 mM of each dNTP, 0.15 mM forward and
35 reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer)

- 78 -

and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were purified (Qiagen PCR purification kit) and direct sequenced.

primers :

(97010709) 5'ATTCTTTTGCATCTTCTTGCGTGCG

(97012802) 5' CGCTCCCCATCAGATGCAGGCCGG

PCR fragment E1.3-3

EST 01222 is homologous but not identical with the 5' end of HU-Unc53/1. A primer pair consisting of one 3' EST 01222 primer and one 5' HU-Unc53/2 primer were used to PCR amplify this fragments. Marathon cDNA from human HeLa S3 was used as template in a PCR. A 50 µl reaction mix contained 1xPCR II buffer (Perkin-Elmer), 1.5 mM MgCl₂, 0.02 mM of each dNTP, 0.15 mM forward and reverse primer, 2.5 U AmpliTaq Gold (Perkin-Elmer) and 1 ml template. The cycling parameters were 5 minutes at 95°C, 35 cycles of 45 seconds at 94°C, 45 seconds at 65°C and 2 minutes at 72°C. The PCR products were sliced out from an agarose gel and purified using a gel extraction kit (Qiagen), one ml hereof was used in a second round PCR using the same conditions as above. The PCR products were analysed on an agarose gel, the fragment of interest was sliced out, purified (Qiagen PCR purification kit) and cloned into pCRIT2.1. The sequence of the insert was determined.

- 79 -

RACE fragments A2.2-2, B2.1-4, D2.1-5; 5'
extension

5 Method as described previously.
gene specific primer (97041701)

5'TATGCTACGGCCACTCATCTCCGTGG
nested gene specific primer (97041702)

10

5'TGTAACCTGAGTTCCCCTTAAACTGG

Marathon cDNA library :

human placenta (fragment A2.1-2)

human HeLa S3 (fragment B2.1-4)

15 human colorectal adenocarcinoma SW480 (fragment
D2.1-5)

Translation-initiation splice variants, fragments
D4.1-1, J4.1.4, G4.1.1, F4.1.2

20

Four different translation initiation splice
variants were detected by 5'RACE.

Method as described previously.
25 gene specific primer (97080803)

5'TCGGTTGTTAGCAGTAGTTGACCCTCC
nested gene specific primer (97080804)

30 5'ACCTGAAAGTCTGGACTGCATTTTCAGC
Marathon cDNA library : human colorectal
adenocarcinoma SW480 (fragment D4.1-1) gene specific
primer (97080801)

35 5'ACAACCTGGATAATCTGGGCCAGGAGG

- 80 -

nested gene specific primer (97080802)

5'TCTTGCTGGAGATCCTTGATGAGACGC

Marathon cDNA library :

5

human melanoma G361 (fragment J4.1.4)

human HeLa S3 (fragment G4.1.1)

human placenta (fragment F4.1.2)

10

DNA sequencing

PCR amplification products and cDNA clones were subcloned either into pBluescript vectors (Stratagene) or in PCR-IIa vector (Invitrogen) and sequenced either manually by the dideoxynucleotide chain termination method with modified T7 DNA polymerase (Sequenase, United States Biochemical) or automatically with an Applied Biosystems 373 DNA sequencer using the fluorescent terminator kit (Perkin Elmer).

20

RNA blots

A Human multiple tissue Northern (MTN-1, Clontech) containing in each lane 2 mg of poly A + RNA from eight different human tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas) and a MTN-II human multiple tissue Northern, containing in each lane 2 mg of poly A + RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral leukocyte, were hybridized according to the manufacturer's instructions and washed out in 0.1xSSC:0.2% SDS at 55°C. Also from Clontech, a poly A + RNA blot from human cancer cell lines (melanoma G361, lung carcinoma A549, colorectal adenocarcinoma SW480, Burkitt's

35

- 81 -

lymphoma Raji Leukemia Molt 4, lymphoblastic leukemia K562, HeLa S3 and promyelocytic leukemia HL60) was tested.

5 Construction of plasmids

Plasmid pCDU2 (Figure 17) was constructed by cloning the 2.8 kb *ApaI*-*NarI* fragment from pTB72, the latter restriction site made blunt with klenow enzyme, into pCDNA3, digested with *EcoRV* and *ApaI*. pCDU2 encodes for the homology blocks A, B, C, D and E. Plasmid pCDU3 (Figure 15) was constructed by cloning the 1.9 kb *ApaI*-*NdeI* fragment from pTB72, the latter restriction site made blunt with Klenow enzyme, into pCDNA3, digested with *EcoRV* and *ApaI*, pCDU3 encodes for the homology blocks C, D and E. Plasmid pCDU4 (Figure 16) was constructed by cloning the 1.4 kb *ApaI*-*StyI* fragment from pTB72, the latter restriction site made blunt with Klenow, into pCDNA3 digested with *EcoRV* and *ApaI*. pCDU4 encodes for the homology block E.

Expression of a domain of the human UNC53 in eukaryotic cells

25

1. pCB201: Equivalent construct of human 1 homologue to expression construct pCDU4 of *C. elegans* unc-53 gene cloned in a eukaryotic His-tag, Xpress Ab tag expression vector.

30

A suitable Bam HI site was engineered on pHH15 open reading frame by amplification with hh15fw primer 5'AGAGCGGATCCATATGCCTCCTTGCCGTCAAGGTG-3' and M13rv primer (5'-cag-gaa-aca-gct-atg-ac-3'). The amplified fragment was then moved to pCDNA3.1.His-A-Vector

35

- 82 -

digested with BamHI and EcoRI. This new plasmid called pCB201 (Figure 13) produces a cDNA which codes for a fusion protein consisting of a 49 amino acid aminoterminal fragment containing an His-tag and also
5 a T7 epitope tag followed by amino acids 1255 to 1627 of the sequence of the human homologue. pCB201 was also checked by sequence and the n was used in stable transfection experiments carried out in N4, MCF7 and NIH3T3 cells.

10

2. pLM5: Equivalent construct of human 1 homologue to expression construct pCDU3 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector.

15

The phage HH3b was linearized using XhoI. A BamHI and XbaaI restriction site were created on the pHH3b open reading frame using U3-Bfw (5'-cca-cac-tag-ggg-atc-cat-gca-aat-gag-g-3') and U-rv (5'-caa-aag-tct-cta-gag-gag-gcc-agt-3') as primers. This
20 amplified fragment was then moved to pBluescript KS, digested with BamHI and XbaI. Sequencing of this plasmid, named pCB300, showed an amino acid change from a serine to an asparagine due to a change from guanine to adenine on the position 4237 of the DNA
25 sequence. This fault was repaired by cloning a 1418 bp fragment of pLM1 (see below) (using NarI and XbaI as enzymes) into pCB300 digested with the same enzymes. The phage HH3b fragment of this plasmid, named pLM6 (fig 53), was then removed using BamHI and
30 XbaI, to pCDNA3.1/HisA digested with the same enzymes. This new plasmid, named pLM5 (fig 52), produces a cDNA which codes for a fusion protein consisting of a 49 amino acids aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by aminoacid 1069 to
35 1627 of the transcript of HU-Unc53/1. Plasmid pLM5 was

- 83 -

checked by sequencing and used on transient and stable transfection experiments carried out in N4 cells. The plasmid pLM1 was created using a PvuII and partial BamHI digested fragment of pHH14-3 and a BamHI and SpeI digested fragment of phage HH3b, cloned into pBluescript KS digested with SmaI and SpeI. The pLM1 contains the full transcript of HU-UNC-53/1 available at this moment (see fig 9).

3.pCB251: Equivalent construct of human 1 homologue to expression construct pCDU2 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector

The phage HH3b was linearized using XhoI. A BamHI and XbaI restriction site were created on the pHH3b open reading frame using U2fw (5'-aag-gga-tga-ttc-ggt-cag-gat-cct-tc-3') and U-rv (5'-caa-aag-tct-cta-gag-gag-gcc-agt-3') as primers. The amplified fragment was then moved to pCR2.1. This plasmid was named pCB250. The pHH3b fragment was removed from pCB250 using BamHI and XbaI and cloned in pCDNA3.1/HisC digested with the same enzymes. This plasmid, named pCB251 (figure 55), was checked by sequencing. pCB251 produces a cDNA which codes for a fusion protein consisting of a 49 amino acid aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by amino acids 828 to 1627 of the partial transcript of HU-Unc53/1. pCB251 was used on transient and stable transfection experiments carried out in N4 cells (see fig 56).

4. pLM3: the partial transcript of HU-Unc531 cloned in an eukaryotic His-tag, Xpress Ab tag expression vector

35

- 84 -

pLM1 was digested with EcoRV and XbaI. This fragment was cloned in pcDNA3.1/HisB, digested with the same enzymes. pLM3 produces a cDNA which codes for a fusion protein consisting of a 49 aminoacid aminoterminal fragment harboring a His-tag and a T7 epitope tag, followed by amino acids 1 to 1627 of the transcript of HU-Unc53/1 available at this moment. pLM3 was used on transient and stable transfection experiments carried out in N4 cells.

5. pLM4: the partial transcript of HU-Unc53/1 cloned in an eukaryotic GFP expression vector

pLM1 was digested with ClaI and XbaI. This fragment was cloned in pEGFP-cl, digested with AccI and XbaI. This plasmid was named pLM4. This plasmid produces a cDNA which codes for a fusion protein consisting of GFP, followed by aminoacid 1 to 1627 of the transcript of HU-Unc53/1. pLM4 was used on transient and stable transfection experiments carried out in N4 cells (see figs 43 and 46).

Stable transfection of MCF-7 cells:

Cells were seeded at a density of 2×10^6 cells in a 75 cm² flask using standard culture medium ((Dubecco's MEM, 450 mg/l glucose, 862 mg/l L-Alanyl-L-Glutamin, 110 mg/l Na-pyruvate; GibcoBRL) supplemented with 10% foetal calf serum (FCS; GibcoBRL), and 100 U/ml penicillin (GibcoBRL) and 100 µg/ml streptomycin). The culture was grown at 37°C in a 10% CO₂ atmosphere, to approximately 70% confluency (approximately 18 hours). The culture medium was removed and 10 ml MEM-HEPES (GibcoBRL) supplemented

- 85 -

with 10% FCS was added to the cells. The culture was further incubated for four hours at 37°C in standard sterile air. DNA-CaCl₂ was meanwhile prepared by mixing 30 µg DNA in 0.1 x TE (1 mM Tris. HCl, pH 7.2, 0.1 mM EDTA, pH 8) and 0.1 ml 1.25 M CaCl₂/HEPES (1.25 M CaCl₂, 0.125 M HEPES; pH 7.05). 0.1 x TE was added to a final volume of 0.5 ml. The DNA-CaCl₂ was added drop by drop to 0.5 ml BS/HEPES (25 mM HEPES, 0.25 M NaCl, 0.01 M KCl, 1.4 mM Na₂HPO₄, 0.01 M glucose, pH 7.05) while pipeting a sterile airflow through the latter solutions. The DNA-Ca₃(PO₄)₂ precipitate was then placed at 37°C for ten minutes. The DNA-Ca₃(PO₄)₂ precipitate was vortexed and added to the cells, together with 100 µl of a 0.01 M chloroquine (Sigma) stock in H₂O. After four hours of incubation at 37°C in sterile standard air, the medium was removed, and the cells were washed with PBS (GibcoBRL). 25 ml of medium was added and the cells were placed at 37°C in a 10% CO₂ atmosphere. After 48 hours of incubation, the cells were harvested, diluted and cultivated under selection (600 µg/ml G418 (Duchefa)) for two weeks prior to clone selection. Mock transfected MCF-7 were positive for the beta-galactosidase transgene. The stability of transfection in MCF-7 was assessed by passaging cells four times in the absence of Geneticin and then re-exposing them to the selector agent. In these experiments, unc-53 or mock transfected cells proliferated, whereas untransfected MCF-7 cells proliferated at a much slower rate.

Stable transfection of N4 neuroblastoma cells

Cells were seeded at a density of 2x10⁶ cells in a 25 cm² flask using standard culture medium ((MEM Rega 3; GibcoBRL) supplemented with 10% FCS, 0.14%

- 86 -

Na₂CO₃,
2 mM glutamine, 100 U/ml penicillin, and 100 µg streptomycin). The culture was grown overnight at 37°C in a 10% CO₂ atmosphere. Transfection mixture was prepared by adding 12 µg DNA in 600 µl opti-
5 mem 1 (GibcoBRL) to 36 µl LipofectAMINE (GibcoBRL) in 600 µl opti-
mem 1. This was done by adding drop by drop the first solution to the second. The mixture was placed for 30 minutes at room temperature, after which
10 1.8 ml of opti-
mem 1 was added. In the meanwhile the cell culture was washed twice with opti-
mem 1, and the 3 ml of transfection mixture was added. The culture was placed at 37°C in sterile standard air. After four hours, 3 ml of normal culture medium was added
15 and the culture was placed at 37°C under 10% of CO₂. 18 hours later, the culture was washed with PBS, and fresh normal culture medium was added. A further 24 hours later, the cells were harvested, diluted and cultured under selection (750 µg/ml G418) for two
20 weeks prior to clone selection.

Fixation of cells for Immunofluorescence

Medium was removed from the 9 cm² wells
25 containing the coverslips. A 4% solution of paraformaldehyde (Sigma) in PHEM (1 g/l glucose, 0.4 g/l KCl, 8 g/l NaCl, 0.06 g/l KH₂PO₄, 0.0475 g/l Na₂HPO₄, 0.35 g/l NaHCO₃, 1.51 g/l PIPES, 0.76 g/l EGTA, 0.19 g/l MgCl₂; pH 6) was added for 30 min at
30 room temperature. The fixative was removed, and the coverslips were washed three times 10 minutes with PHEM. The coverslips were then placed in PHEM, containing 0.5% Triton-X100 (Serva) for 30 minutes, after which the slide was washed again for three times
35 10 minutes with PHEM. The coverslips were then placed

- 87 -

under PBS (0.14 M NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.3) containing 0.2% Tween (Sigma) for at least one hour at 4°C

5 Immunofluorescence staining

 The coverslips were inverted on 35 µl of appropriately diluted antibody, being YL 1/2 for tubulin and/or mab 16-48-2 monoclonal or anti-UNC53
10 (gp48) polyclonal antibody for UNC53. The slides were placed at 4°C for at least 18 hours. Excess of primary antibody was then removed by washes of three times ten minutes in PBS-Tween. The slides were then treated with secondary antibody in the same way as for
15 the primary antibody. F-actin was labelled by including TRITC- or FITC coupled phalloidine to the incubation buffer. The inverted slides on the secondary antibody were left at room temperature for approximately one hour. Slides were then washed again
20 for three times ten minutes with PBS-Tween and once with PBS. The coverslips were mounted on slides with the medium described by Herzog et al. (Cell Biology: a laboratory handbook, 1994, Academic Press, 355-360). After at least two hours, slides were ready for
25 analysis.

Time lapse analysis

 Analysis of the behaviour and movement of growing
30 cell cultures was done by placing a non-confluent culture under a phase contrast microscope equipped with a temperature controlled stage (37°C). Images were recorded using a CCD camera (COHU 4912) coupled to a SCION LG3 framegrabber in a Macintosh ppc 8100
35 running NIH image version 1.60. Images were recorded

- 88 -

at time intervals, varying from 15 sec to 1 min. for half an hour to two hours. Image enhancement and playback was done in NIH image.

5 Phagokinesis

A variety of cell types were shown to migrate over colloidal gold coated culture plastic or glass and displace or phagocytose the gold lawn on their way while locomoting. The track left bare is a qualitative and quantitative measure of cell motility and/or locomotion. The basic methods have been described in detail elsewhere (Albrecht-Buehler, 1977, Cell, 11: 395, Zetter, 1980; Nature, 285: 41; O'Keefe et al., 1983; J. Invest. Dermatol., 85: 130). Culture plates were gelatin and gold coated as described by Albrecht-Buehler (1977). Unc-53 and mock transfected MCF-7 were seeded in plates at low density and allowed to adhere to the plate and to locomote overnight. Cells were chemically fixed to the plate, washed and air-dried. Images of the gold lawns were captured using automated videomicroscopy; composite images of the wells were generated and single-cell phagokinesis tracks were measured using a home-made routine in SCIL™ software.

C. elegans-UNC-53 preferentially binds microtubule plus ends or GTP-tubulin

30 1. Cloning of C.elegans cDNA in pEGFP-C1 and construction of C-terminal deletion derivatives.

a) Constructing a GFP-Unc53 N-terminal fusion:

35 A PCR experiment was performed under standard conditions, using pTB72 as template and cp17

- 89 -

(ata gcc aga tct acg tca aat gta gaa ttg) and cp18 (ttt aga aac cgc ggg tgg) as primers. The resulting 0.4 kb fragment, coding for the N-terminal fragment of C.elegans Unc 53 was cloned in vector pCR2.1 (original
5 TA cloning kit, Invitrogen), resulting in plasmid pTA1718. The 0.4 kb fragment was isolated as a BglIII-SacII fragment and cloned in pEGFP-C1 (Clonetech) digested with the same enzymes. The resulting plasmid was designated pEGFPsac (Figure 29). pEGFPsac encodes
10 the N-terminal 13 aa of C.e.Unc53 in fusion with GFP.

b) Construction of a GFP-C.e. Unc53 full length fusion:

Plasmid pTB72 (shown in Figure 1) was digested with restriction enzymes SacII and ApaI. The
15 resulting 4.5 kb cDNA fragment, encoding for the C-terminal fragment of C.elegans Unc53 was cloned in plasmid pEGFPsac (Figure 29), digested with the same enzymes, resulting in plasmid pEGFP72 (Figure 30). Plasmid pEGFP encodes GFP in fusion with the full
20 length C.e. Unc53.

c) Construction of N-terminal deletions of GFP-C.elegans UNC-53 fusion protein, other than pEGFPsac:

pEGFP72 was digested with SmaI. The resulting 7.0 kb fragment was religated and
25 transformed in E.coli, resulting in plasmid pEGFPsma (Figure 31). This plasmid codes for the first 760 aa of the Ce-UNC-53 in fusion GFP.

pEGFP72 was digested with restriction enzymes Ec1136II and SmaI, the resulting plasmid after
30 ligation and transformation in E.coli of the 6.7 kb fragment was designated pEGFPec1 (Figure 32). This plasmid codes for the N-terminal 670 aa of the C.e. Unc53 in fusion with GFP. pEGFP72 was further digested with SmaI and XbaI. The latter site was made
35 blunt with Klenow polymerase. The resulting fragment

- 90 -

of 5.4 kb was religated and transformed in E.coli. The resulting plasmid was designated pEGFxba (Figure 33). This plasmid codes for the N-terminal 256 aa of C.elegans Unc53 in fusion with GFP.

5

2. Constructing a hul-UNC-53-GFP fusion, and a deletion derivative

The 5.4 kb hul-unc53 fragment was isolated as ClaI-XbaI fragment from pLM1 (Figure 54), and cloned in pEGFP-C1 digested with AccI and XbaI. pEGFP-C1 was isolated from E.coli GM41 (Hfa H, dam-3, thi-1, rel-1). This makes the XbaI restriction site available for restriction digest. The resulting plasmid was designated pLM4 (Figure 34).

15

3. Visualisation of GFP fluorescence in N4 cells

N4 neuroblastoma lines were seeded in Lab Tek chambered coverglass (Nalge Nunc International) and transfected using lipofectAMINE (GibcoBRL). After 18 hours, the chambered coverglasses were placed on an inverted microscope, and GFP fluorescence could be visualised.

25

4. Staining GFP fusion expressing cells with antibodies

Transfection with the GFP fusion constructed was also performed on coverglasses in a 6-well plate. After paraformaldehyde or methanol-acetone fixation, cells could be stained for actin cytoskeleton with TRITC-phalloidine, for hul-unc53 with sera 28.1 and for tubuline with YL1/2 antibody. Visualisation was then

35

- 91 -

performed on a axioplan (Zeiss microscope).

5 Methods of Producing and Observing the Effects of
 A Chimeric unc-53 Gene

 1. Definition of a promoter region in the unc-53 C.elegans gene:

10 The genomic region from the position 15621 to
 18415 in the C.elegans unc-53 gene, called promoter A,
 was cloned and fused to the cDNA of the GFP gene
 (clone pA/GFP, or pNP10)(cf. fig.51). This construct
 is injected into wild type worms (N2). Transgenic
15 line express GFP in different neurones: the two pairs
 of pioneering neurones PVP and PVQ, both BDU neurones,
 both ALN and PLN neurones, both PDE neurones, both PVM
 neurones, and 4 vulval cells. Expression begins in
 early embryogenesis, when the axons of those neurones
20 grow out.

 2. Mutant Phenotype in Unc-53(n152) alleles:

25 In wild type worms (N2), the two pairs of ALN and
 PLN neurones each send an axon in a straight line
 longitudinally from the tail to the head (see
 fig.50a). In unc-53(n152) alleles, the axons are
 shorter and often branch in a dorso ventral direction
30 (see fig.50b). The neurones are visualised with the
 construct pA/GFP, injected in unc-53(n152) worms.

 3. The minigene pA/unc-53 rescues the
 elongation defect of ALN and PLN neurones:
35

- 92 -

The promoter A from the *C.elegans* unc-53 gene was fused to the cDNA of the *C.elegans* unc-53 gene (clone pA/unc53, or pNP9). This construct was injected in unc-53(n152) mutant worms, together with the pA/GFP construct described above to visualise the ALN and PLN neurones. The elongation defects of those neurones in the unc-53 mutant are almost completely restored by the expression of the unc-53 cDNA express under the promoter A (see figs. 50 and 51b).

4. Domain swap between the *C.elegans* and human unc-53 gene:

To test whether the vertebrate and worm members of the unc-53 family are functionally equivalent, we tested the ability of the human gene to rescue the mutant phenotype in the worm. We replaced the carboxyterminal predicted nucleotide binding domain (NTPase) of the worm protein with the homologous fragment of the human 1 gene.

The clone pA/unc-53 was deleted of the *C.elegans* NTPase domain, from the HpaI site, position 29800 on the genomic of unc-53, and replaced by the equivalent domain of the human-1 gene (unc-53H1) (see fig. 51). The resulting clone is named pA/unc-53H1. When this clone is injected to unc-53(n152) mutants, the transgenic worms show a significant but incomplete rescue of the defect in the elongation of the ALN and PLN neurones (see fig. 51b). The axons are longer, often elongated until the region of the vulva in a straight line, without branching dorsally anymore. This result shows that a NTPase region of the human unc-53 homologue can functionally replace the NTPase region of the *C.elegans* worm.

- 93 -

The degree of rescue was analyzed quantitatively and summarized in Figure 51b:

The four strains compared are:

wt; un-53(n152); unc-53(n152), pA/unc-53; unc-

5 53(152), pA/unc-53-H1.

The various phenotypes observed are brought together in three large classes:

<<wild type>> the axon is straight, unbranched and migrates into the head;

10 <<vulva>> the axon is straight, unbranched and stops in the vulva region;

<<mutant>> the axon is short, does not reach the vulva region and has collateral branches.

The figures are indicated as a percentage. The number
15 of axons observed is indicated in the following column.

The data clearly show demonstrate conclude that the nematode/human chimera minigene pA/unc-53-H1 partly rescues the defects of the axonal migration of
20 the ALN and PLN neurones and demonstrate conservation of function of this domain between man and worm. The transgenic lines provide a functional screening assay for the motility function of at least part of the human UNC-53 gene.

25

II. Materials and methods

1 - Cloning:

30 a) pAB/GFP (pNP3 - Figure 27)

The gene of GFP has been amplified by PCR with cpn3 oligo-nucleotides

"acattaagcttcgtacgcttgagggtaccg" and Cpn5

35 "gaaaggatccgtacgataaggtattttgtgtcgg" on the plasmid pPD95.75(Figure 59) so as to be inserted at the 5'

- 94 -

position in fusion into the exon 12 of the unc-53 gene at a single restriction site SphI and contains its stop codon at 3' plus one polyadenylation site. The PCR amplification product is directed by HindIII and BamHI, sites which are contained respectively in the cpn3 and cpn5 oligonucleotides and sub-cloned in the pBS vector (clone pNP2). The GFP is then excised from the pNP2 clone at the site SphI and integrated into the X16 clone (Figure 60) originating from sub-cloning of the lambda phage S4 digested by XhoI. The X16 clone containing the genomic sequence of unc-53 from the position 16621 to the position 24891 cloned in the site XhoI of pBS.

b) pAB/unc-53 (pNP8 - Figure 35)

The promoter region AB of the X16 clone (between PstI and SphI) has been inserted in the clone pTB115 (Figure 58) in which the region between the sites PstI and SphI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

c) pA/GFP (pNP10 - Figure 56)

The promoter region A has come from the X16 clone between the sites PstI and NheI and integrated in the vector pPD95.75 containing the GFP in the sites PstI and XbaI.

d) pA/unc-53 (pNP9 - Figure 44)

The promoter region A has come from the X16 clone between the sites PstI and BstXI and is integrated into the clone pTB115 in which the region between the sites PstI and BstXI, containing the promoter of the gene mec-7 and the start of the gene unc-53, has been removed.

e) pA/unc-53 -H1 (pCB501 - Figure 57)

- 95 -

The clone pA/unc-53 (pNP9) has been deleted from the region 3' of the gene unc-53 of the nematode between the sites HpaI and NcoI. The 3' region of the Hlunc-53 gene has been amplified by PCR with the oligonucleotides U4Afw (5'-gca-cat-cgt-taa-cgg-gga-
5 ctt-gaa-gc-3') and Urv (5'-caa-aag-tct-cta-gag-gcc-agt-3') and digested with HpaI and XbaI. After a filling stage with T4 polymerase, the ligation is effected with a complete end.

10

2-Injection

Conventional injection techniques are used (Fire A., 1986, Mello G, et al, 1991, journal Mello G. and Fire A., 1995). Young hermaphrodite adults are injected in their two syncytial gonads. The DNA used is prepared in standard manner (Qiagen) followed by precipitation with lithium chloride. After an extensive rinsing stage to eliminate all the salts, the DNA is
20 resuspended in water. The injection solution contains the different DNAs at a concentration of 100 ng/μl in an injection buffer. The plasmid pRF4 containing the dominant allele su 1006 of the gene rol-6 (Kramer J. et al, 1990, Mello C. et al, 1991) is used as a
25 transformation co-marker. The descendants of roller phenotype of the hermaphrodite injected are isolated. Approximately 10 % of these transformants will yield a stable strain, in which the different DNAs injected are associated to form a mini-chromosome which will
30 segregate as unstable extrachromosomal arrays. All the transgenic strains obtained were tested by PCR for the presence of the DNA injected, using a specific primer of the vector and a primer in the gene (results not shown).

35

3. Microscopy

- 96 -

The nematodes are observed under a ZEISS Axioplan microscope provided with Nomarski lenses, with 40X Neofluar, 63X Plan-Apochromat, 100X Plan-Apochromat objective lenses. For fluorescence observation the luminous source is a mercury bulb. Different ZEISS filters are used:

- for observation under GFP fluorescence, FITC filter: blue excitation line at 588 nm, emission through a 515-565 nm band-pass filter;

- for observation of the antibody labelling with a secondary antibody coupled to the TRITC: excitation through a 546 nm band-pass filter, emission through a 590 nm long-pass filter.

The image acquisition is effected by means of a CCD camera and and NIH image program using a Machintosh computer. The images are processed using the Adobe Photoshop program.

Sequence Listing

The following sequences are referred to in the specification:

5

Sequence ID No 1 is an amino acid sequence of human homologue 1 of UNC-53 protein illustrated in Figure 9b.

10 Sequence ID No 2 is an amino acid sequence of human homologue 2 of UNC-53 protein illustrated in figure 11d.

Sequence ID No 3 is a nucleic acid sequence of the hu-1-unc-53 gene illustrated in Figure 9b.

15 Sequence ID No 4 is a nucleic acid sequence of the hu-2-unc-53 gene illustrated in Figure 11d.

Sequence ID No 5 is a nucleotide sequence of Phage Lamda Clone 3b deposited under Accession No LMBP 3595.illustrated in Figure 9.

20 Sequence ID No 6 is a nucleotide sequence of plasmid pLM1 deposited under Accession No LMBP 3762 and illustrated in fig 54.

Sequence ID No 7 is a nucleotide sequence of plasmid pLM4 deposited under Accession No 3763 and illustrated in fig 34.

25 Sequence ID No 8 is a nucleotide sequence of plasmid pEGFP72 deposited under LMBP Accession No 3764 and illustrated in fig 30.

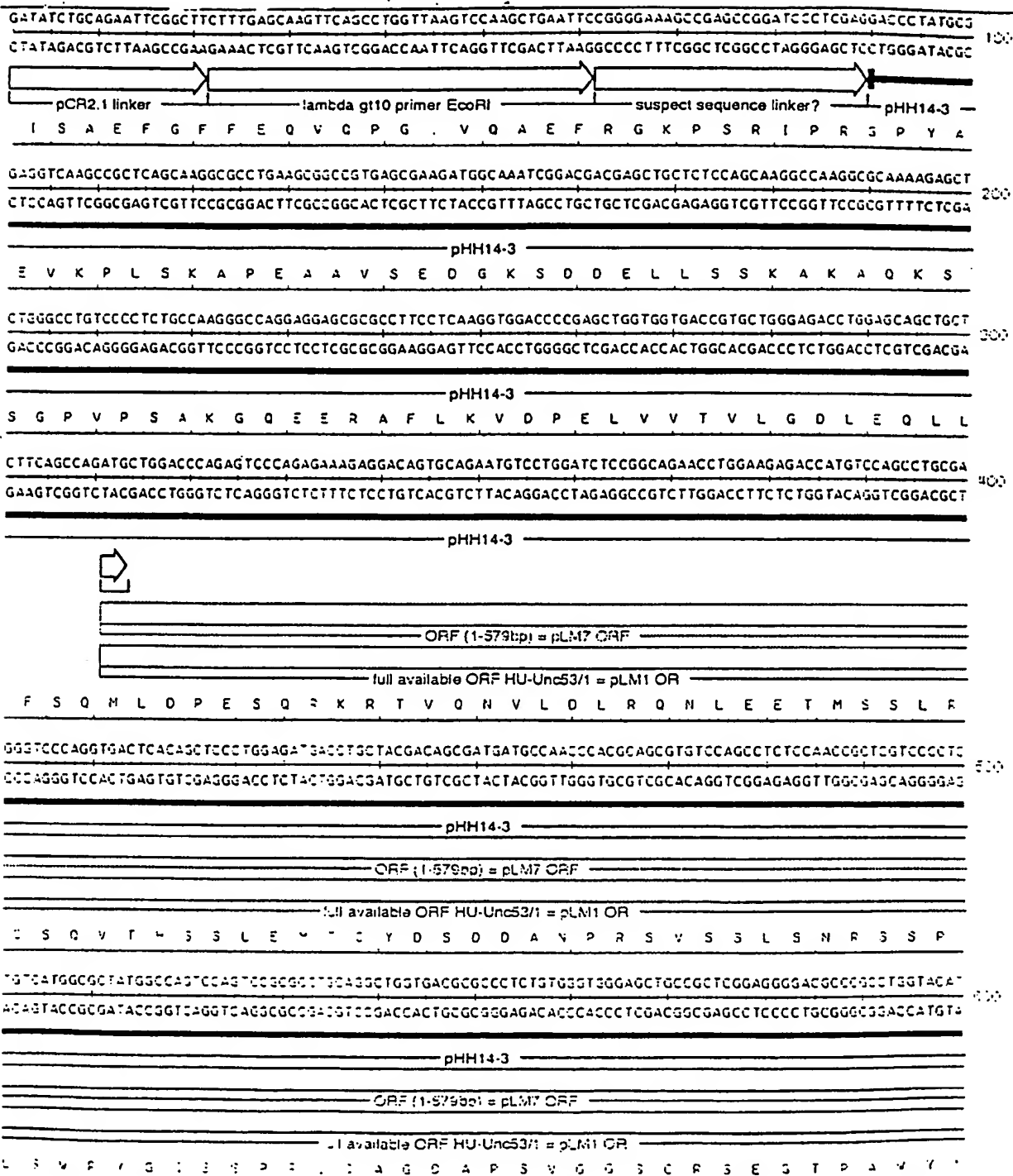
30 Sequence ID No 9 is a nucleotide sequence of plasmid pCB501 deposited under Accession No 3765 and illustrated in fig 57.

Sequence ID No 10 is a nucleotide sequence of plasmid pCB201 deposited under Accession No. LMBP 3594.

SEQ ID NO. 1

AMINO ACID SEQUENCE OF
HU-UNC-53/1 PROTEIN

Page 1



Tuesday, 18 November 1997 10:33
fig. Hu-Unc53/1 seq (1 > 6013) Site and Sequence

F. 9 9

Page 2

GCACGGCGAACGGGCCCTACTCCACACCATGCCATGCCAGCCCCAGCAAGCTCAGCCATATC TCCCGCCTGGAGCTGGTCGAATCCCTGGACTCG 700
CGTGCCGCTTGCCCGGGTGATGAGGGTGTTGACGGGTACGGCTCGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC
pHH14-3
ORF (1-573bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L D S
GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCAC TACCGGCTGGGATGAAAGCA 800
CTACTCCACCTGGAGTTACAGGCCGATGTACTCGCTGTCACCTGGAGTACCCGTTCTGGTACTTGCCCTCTACTACTGTAGTGATGGCCGACCTTACTTTCTGT
pHH14-3
ORF (1-573bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D E S
GCTCCATCAGTAGTGGACTCAGCGATGCCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 800
CGAGGTAGTCATCAGTACGTCGCTACGGAGTCTGTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTTCATGAGGGTGAGC
pHH14-3
pCB212
ORF (1-573bp) = pLM7 ORF
full available ORF HU-Unc53/1 = pLM1 OR
S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P T A
TTCTCGCAGGAACCTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCC 900
AAGAGCGTCTTGAATGTATCAGGATGCGGTCTCTAGTCTCTCGCGAGTGACCGTCTTACCCGACTCGACCAATCAGTTAGTCTCCCTCTTTCTGG
pHH14-3
pCB212
full available ORF HU-Unc53/1 = pLM1 OR
E R R N S T I V L R T C S E K R S L A E S G L S V F S E S E E V A
CCTAAAAAATGGAGTACGACAGTGGTAGCCTGAAGA "GGAACCTGGGACTTCTAAGTGGCGGAGGAGCGGCTGAGAGCTGTGATGATTTCACCAAGG 1000
GGATTTTITGACCTCATGCTGTACCATCGGAC "TCTACCTGGACCTTGAAGATTCACCGCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCT
pHH14-3
pCB212
full available ORF HU-Unc53/1 = pLM1 OR
L E A D S G S L V E P G T S K V R R E D P E S C D D S S I

Tuesday, 18 November 1997 10:33

Fig 9

Page 3

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTGGAGAAC TGAAAAAGCCATCAGCCTGGGCCACCC TGGTTCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAAC TTCCCCATCAGTCACACAGC
CCCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTC TTCCCGTTCTGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGCG 1200

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGGCCTCAAAGTCGAGGCAAACTGAGGGCAAGCTTACAGACAAGGGTAAGCTTGCAAGTGAAGAATAC TGGGCTCCACGC TCCTCTCTGAT
GGTCTCACGGGAGTTTCAGCGTCCGTTTGAC TCCCTTCGATGTCTGT TCCCATTCGAACGTCAC TTCTATGACCCGAGGTTGCGAGGAGGAGACTA 1300

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCC TCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATTC TCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT 1400

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTGATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCGTTGAGAGTCGTTC TAGGTCTTCAGGAGTCCGTAGGGACAGTTTCGGTCAATTACCCGCGTTCG 1500

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R I T

TAGCTTAGATGTTTCCAACAGTSCAGAGCCAGGATTCCTSGCTCCTGGAGCCCGTTC TAACA TCCAGTACCGCAGCCTGCCCGGCCAGCC AAGTCAAGT
ATCGAATCTACAAAGTTGTCAGTCTCGGTCCTAAGTACCGAGGACC TCGGGCAAGATTGTAGGTATGTCGTCGGACGGGGCCGGTCGGTTCAGTTCA 1600

pHH14-3

pCB212

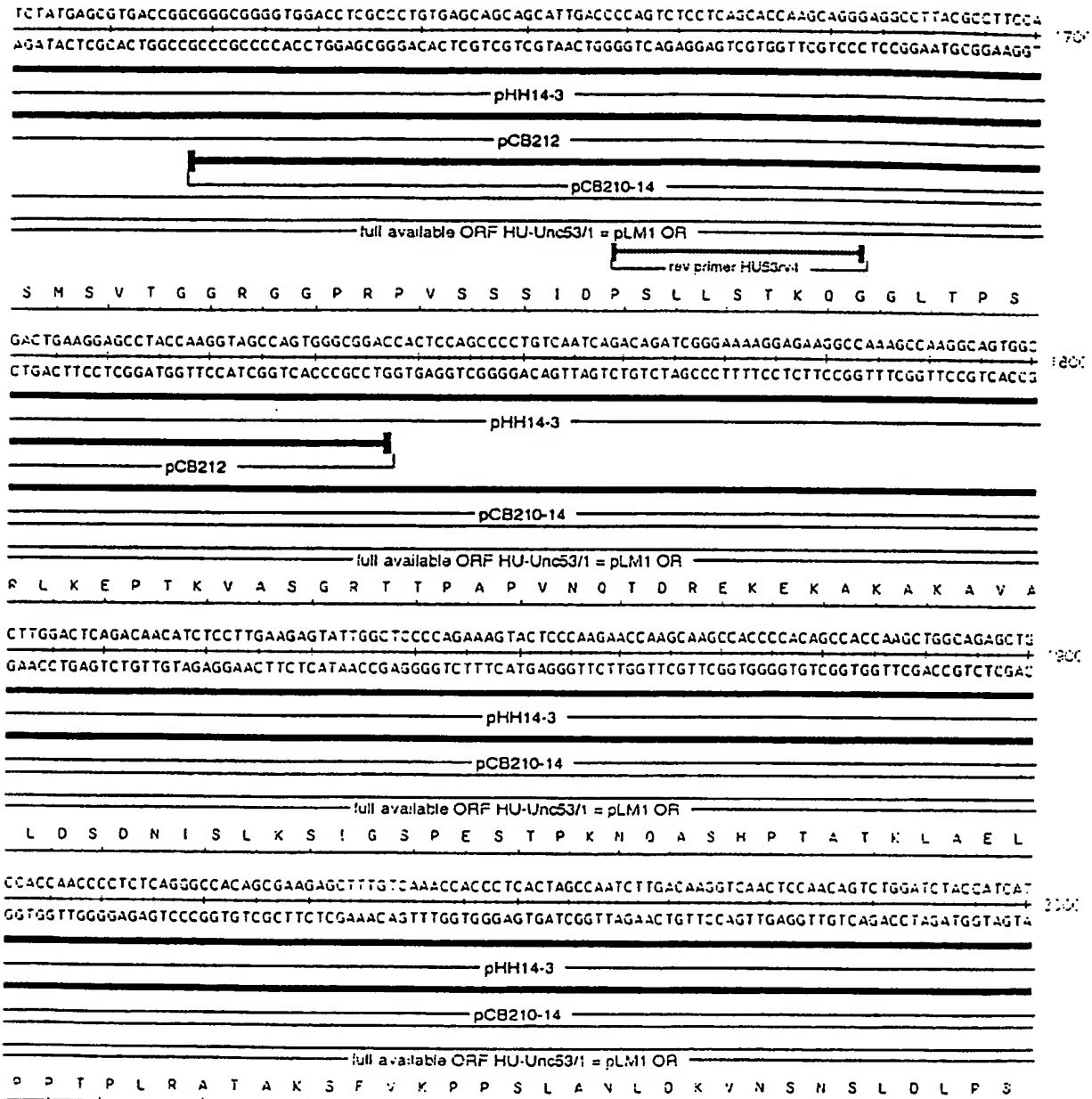
full available ORF HU-Unc53/1 = pLM1 OR

I L D V S N S A E P G F L A P G A R S N I O Y R S L P R P A N S S

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

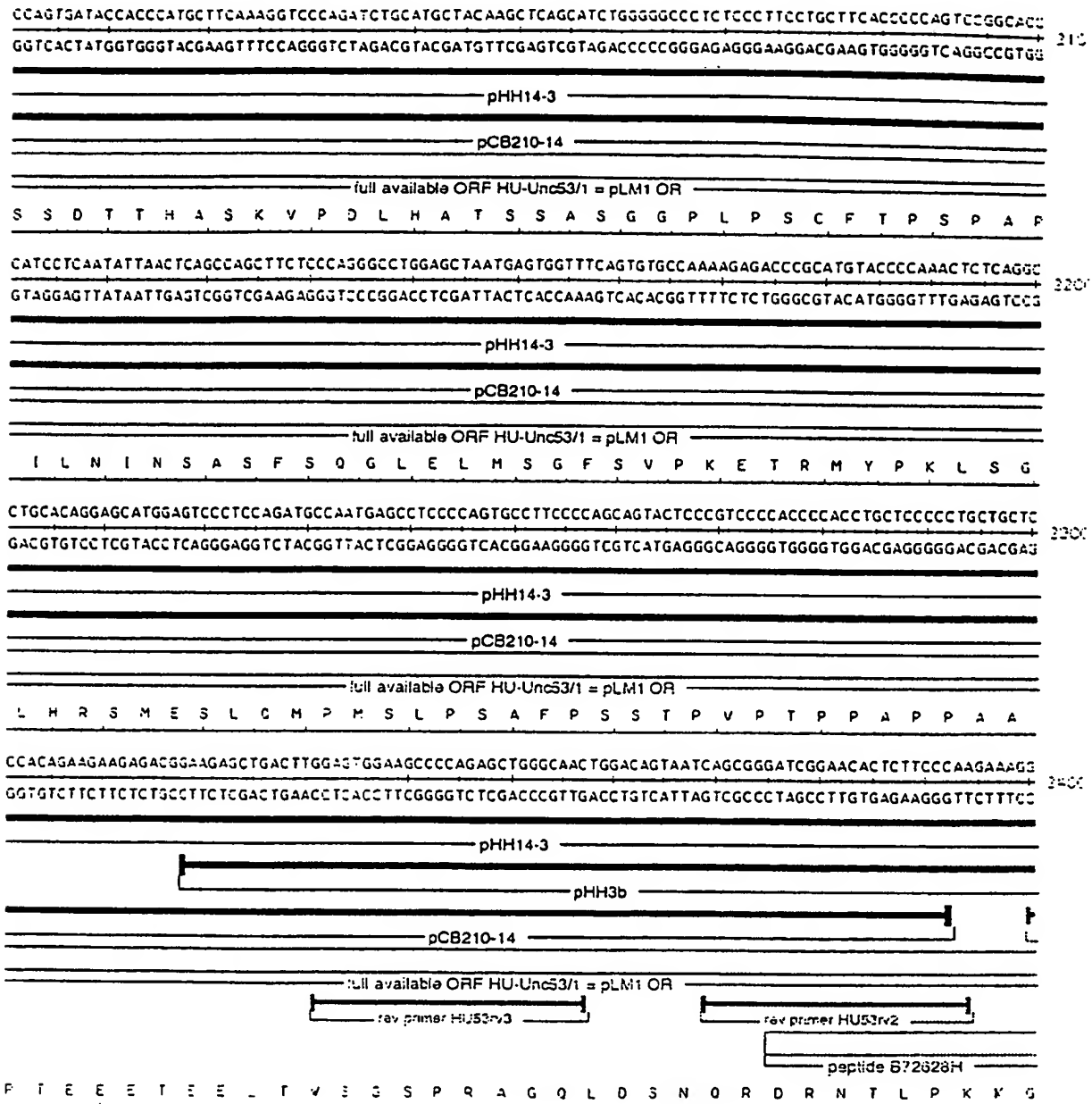
Page 4



Tuesday, 18 November 1997 10:33
fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 5



Tuesday, 18 November 1997 10:33

fig HU-Unc53/1 seq (1 > 6013) Site and Sequence

Page 6

GC TCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGGGACATTCCCATACCATTTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTCGAAGTCAGGGTCTCTCTCTGGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAA 250

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACCTTCCCATGTCTCTGAGTGCAAGGGGCCAACTTACCAACATAGTGAGTCCCCTGCGGCCACCAAGCAAGCAATCAGCCGCTCCAACT
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACCGGGTGGTGGGTTCTTAGTGGGCGAGGTTGT 260

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A X G Q L T N I V S P T A A T T P R I T R S H

GCATCCCCACCCACGAGGGCGGCTTCGAGCTGTACAGCGGCTCCCAATGGGGAGCACCCGTGTCCTGGCCGAGAGACCAAGGGAATGATTGGGTCAGG
CGTAGGGGTGGGTGCTCCGCGGAAGCTCGACATGTGCGCGAGGGTTACCCCTCGTGGGACAGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC 270

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S O M G S T L S L A E R P K G M I R S G

ATCCTCCGAGACCCACGGACGATGTTACGGGCTCAGTCTCTCTCCCTGGCCTCCAGTGCTTCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCT
TAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTACGACAGGGACCGGAGGTACGGAGGAGGTGATGAGGAGTCGACTCCTCTCTTACGTAGGA 280

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M Q S

GACCAATCCGGAAGCTTCGTAGGGAAC TGAATCATCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT
CTCGTTTAGGCTTCGAAGCATCCCTTGACCTTAGTAGSSTCTTTTACCAGGTGGAAC TSCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA 290

U2 ORF = pC0351 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I R K L R P E L E S S G E K V A T L T S C L S A N A N L V A A

Fig 9

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 7

TTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTCGAGAGAAACCATAGA
AACTCGTCTCGGACCACCTTATACTGTAGGGCGGACGCTGTGGACCGCTCTCGCCGGCTCCTCTTCCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT 300

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTTCTGAAGAAAAGAACTCTGAGGCCAGGCGAGTCATTCAGGGAGCCCTTAATGCCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAC
GAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTCTTGAAGCCTAGTCTCTGTTTTG 310

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCTAGTGTCTGGTAAGGTCGTAGCCGTCGTCTTCTACGACTACGCTTTTCTCTTTTTTCTCAACCC 320

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K X K K K S V

TCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCACTATAAAAAAGGGGCCAAGTCAGCTTCTCTCATCTCGGATATAGAGGAGATTGCTACACCCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCTATTTTTCCTCCCGGTTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAAGSATGTGGGCT 330

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTCAGCCCCCTCATCCCCCAAACTACAGCATGGTCTACAGAGACTGCTTCACCCCTCCATCAAGTCTCTCACCTTGTCTCTCGGTGGGCACTGATGTC
GAGAACTCGGGGGAGTAGGGGGTTTGATGTCTCTACCAAGATGCTCTGACGAAGTGGGAGGTAGTTCAAGAGGTGGAAACAGGAGGCCACCCCTGACTACAG 340

U2 ORF = pCB251 ORF

PHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q F G S T E T A S P S I K S S T L S S V S T D V

Tuesday, 18 November 1997 10:33

fir HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 5

ACCC .GCCCTGCTACCCAGCCCCACACTAGGCTGTTCATGCAATGAGGAGGAGGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGC 3500
TGGCTCCCGGGACGAGTGGGTCTGGGGGGTGTGATCCSACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCTTCCATAGCCTCGACGCGAGACTCG

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTT 3600
ATACCCTCTTCTTACTTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAA

U2 ORF = pCB251 ORF

pHH3b

peptide B72627H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTA 3700
CCTCCACCTGGACGACTTTCGTCCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGSGACCTAGTAGACGTAA

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCACGCCGCTCCCTAGGCCCTGSCACTCACCATTCTTGGGCCCCAGTCTTGACAGACACAGACCTGTCAACCATGGATGGCATCAGTACTTGTG 3800
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCCGGGTTCAGAAGCTCTGTGTCGGACAGTGGGTACCTACCGTAGTCATGAACAC

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T H S F G P S L A C T D L S P M D G I S T C

Tuesday, 18 November 1997 10:34
fig. 9 HU-Unc53/1 seq (1 > 6013) Site and Sequence

Page 9

GTCCAAAGGAGGAAGTGACCCCTCCGGGTGGTGGTGAAGATGCCCCCGCAGCACATCATCAAAAGGGGACTTGAAAGCAGCAAGGAATTCTTCCTGGGCTGTAG
CAGGTTTCCTCCTTCACCTGGGAGGCCACCAACCACTCTACGGGGGCGTCGTGTAGTAGTTTCCCTGAACCTCGTCGTCTCTAAGAAGGACCCGACATC

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P O H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCAGGTGTTCAAGGACTATATTCTAAATGGACCCAGCCTCTACCTGGGA
GTTCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCTGATATAAAGATTYTACCTGGGTGGGAGATGGGACCT

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M O P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCAGCTGAAACGAGTGTGGATGCAGAGCCCCCGAGATGCCTCTTGGCGTCGAGGTGTCAATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGCTGCACCTTGTCTCACAACCTACGTCCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

AGATATCAGTCTCCCTCAAAGGCTCTGAAGGAGAAATGGCTCGACAGCCTGGTGTTCGAGAGCGTGAATCCCCAAGCCGATGATGCAGCACTACATAAGGCT
TGATATAGTCAGAGGGAGTTTCAGACCTCTCTCTTTACGACGCTGTCGGACCAACAAGCTCTCGGACAGGGGTTTCGGCTACTACGTCGTGATGTAATCGGA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

H I S V S L Y G L E E C V U S L V F E T L : P A P M M Q H I S L

Tuesday, 18 November 1997 10:34

fig. 9 HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 10

CCCTCGAGGACCGGCGCCTCGTCCCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGG
GGACGACTCGTGGCCGCGGAGCAGGAGAGCCCCGGGTCGCCGTGCCCGTCTTGGATGGACTGGTTAGCGAACC GGCTCATGGACCACTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACAGATAGACC
GCACCTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E V T E S I V S T F N M H Q Q S C K D L Q L Y L S N L A H Q I D

GGGAAACAGGAATTGGGGAATGTGCCCTGGTGATTCTATTTGATGACCTGAGTGAAGCAGGC TCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCA
CCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCGGAGGTAGTCACTCAACCACTTACCCTGGGAGTGGAGCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E T G I G D V P L L L L D D L S E A G S S E L V N G A L T C I

Tuesday, 18 November 1997 10:34
fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 11

G T A T C A . A A A T G T C C C T A T A T T A T A G G T A C C A C C A A T C A G C C T G T A A A A A T S A C A C C C A A C C A T G G C T T G C A C T T G A G C T T C A G G A T G T T G A C C T T C T C C
C A T A G T A T T T A C A G G A T A T A A T A T C C A T G G T G G T T A G T C G G A C A T T T T T A C T G T G G G T T G G T A C C G A A C G T G A A C T C G A A G T C C T A C A A C T G A A A G A G G

U2 ORF = pCS251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72529H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

A A C A A C G T G G A G C C A G C C A A T G G C T T C C T G G T T C G T T A C C T G A G G A G G A A G C T G G T A G A G T C A G A C A G C G A C A T C A A T G C C A A C A A G G A A G A G C T G C T T C
T G T T G C A C C T C G G T C G G T A C C G A A G G A C C A A G C A A T G G A C T C C T C C T T C G A C C A T C T A G T C T G T C G C T G T A G T T A C G G T T G T T C C T T C T C G A C G A A G

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

Y N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

G G S T G C T C G A C T G G G T A C C C A A G C T G T G G T A T C A T C T C C A C A C C T T C C T T G A G A A G C A G C A C C T C A G A C T T C C T C A T C G G C C C T T G C T T C T T T C T G T C
C C C A C G A C T G A C C C A T G G T T C G A C A C C A T A G T A G A G G T G T G G A A G G A A C T C T T C G T G T C T G G A G T C T G A A G G A G T A G C C G G G A A C G A A G A A G A C A G

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L D V V F K L V Y - L F T F L E K H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34
fig. 1J-Unc53/1 seq (1 > 6013) Site and Sequence

Fig.

Page 12

GTGTCCCATTTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACCTGTGGAACAACCTCTATCATTTCCCTATCTACAGGAAGGAGCCAAAGGATGGGATAGAG
CAGAGGGTAAAGCGTAACCTCTGAAGGCCGGACCAAGTAACCTGGACACCTTGTGTGAGATAGTAAGGGATAGATGTCCCTTCCCTCGGTTCCCTACCTTATTC

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCASTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACC AATCAAAGCTGTACCAAC
CAGGTACCTGTCTTTTCGACGAACCTCTCGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V F G C K A A V E D P V E V V R D T L P V P S A Q Q D G S X L Y H

TNCCGCCACCCACCGTGGGCCCTCACAGCATTGCTCACTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCCTCTGA
ACGGGGGTGGGTGGCACCAGGAGTGTCTGAACGGASTGGAGGGCTCTTATCTGTGAGTTTCTGTCTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L P P P T V G P H S I A S P P E O R T V K O S T P S S L D S C F L Y

Tuesday, 18 November 1997 10:34
file -lu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page i 3

GGCCATGCTGCTGAAACCTCAAGAAGCTGCCAACTGATATGAGTCTCCAGATCGAGAAACCACTCTGGACCCCAACCTTCAGGCCAACATTTAAGGSTTC
CCTGGTACGACGACTTTGAAGTCTTCGACGGTGTATGTAATCAGAGGCTTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTGTGAAATTTCCCAAG

U2 ORF = pCB251 ORF

- pHH3b

- U4 ORF = pCB201 ORF

- full available ORF HU-Unc53/1 = pLM1 CR

- U3 ORF = pLM5 ORF

- pHH15

- pupilo 372825H

A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L . G F

GCAATCAC TGTACCCCCGGAGCAGAACGCTGGCATCAGCTATCTTAGCTCCCTCCTCTCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCA
CCSTTAGTGACAGTGGGGGCC TGCTGCTTTCGCACCCTAGTCGATAGAATCGAGGAGGAGAGGAGAGGAAGTCTCTGTACCCAGAGGCTCGGGGTC

- pHH3b

- pHH15

G N H C H P R T A E R W H Q L S . L L L S P L L F Q S T G S P A P

[illegible]

- pHH3b

pH 15

U S E N S G G G D E R S S T G S W C C T F E N F L G R N G G V A F G

GACCTGTGCCCCCAAACACATTTACTGGCCTCTCTTAATGACTTTGGGGAAAGATGATCTGGGCTTTCCCTTGACTTCTTGTTTCAATTACAAAC
 CTGGAACACGGGGGATTTGTGTAATGACCGGAAGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGGAACGGAAGAACAAAGTTAATGTTT

- pHH3b

- pHH15

A L C P L N T F T G L L . . L W G X D O S G S F P . L L V S I T M

TCCTGGGCTTCTCGGGAGGGGTTACAGAAACA¹CTAAACACGCGACGAGTCTCTAAATGATCTCACAAGCAACCCCTGAGAGAGACAGTCTTGTGAGGG
AGGACCCCGAAAGACCCCTCCCAAGTCTTTTGTACTTTTGTGACGTCGTCAGGATTACTAAGAGTGTTCGTGGGACTCTCTCTGTCTAGAACACTCC

- pHH3b

- pHH15

W A F W G G V C D T I : . H C S S S . M I . - S V P E P O S L V R

Tuesday, 18 November 1997 10:34
fig 1Hu-Unc53/1 seq (1 > 6013) Site and Sequence

fig. 1 Ru-Unc53/1 seq (1 > 6013). Site and Sequence

F. 9.

Page 14

AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTCACAGACCCCTCCCAATTCCATCACCAC TGCCAAACAATCTCCCCAGAGATCTGGCTGGAGC
TCTAGACCCCCCTCCGTCCTCGAGGAGTCTAAAGAGTGTC TGGGAAGGGTTAAGGTAGTGGTGACGGT TGTGTGAGGAGGGGGTCTCTAGACCGACCTG 570

-pHH15

E I W G R O E A P O I F S Q T L P N S I T T A N N S S P R O L A G A

CCAGAAAAGAAGCATGTGGTTTAAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAACAAAAACAAGCAAAACAACAAAAACAATGGAA
GCTCTTTTCTTGTCACACAAATTTTACAAATTTAGTTAGACATTTCCATTTTACTTTTTGTTTTGTTGCGTTTGTTTGTTTGTGTACCTTT 520
O K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q W I

AGATGAAGCTGGAGAGAGAGGAACCAAGTTGCCAAGGTAGAGAGCTGCCCGCTCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACCT
TCTACTTCGACCTCTCTCTCTCTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCTGTAGTGTGTCTGCCGACGGTTGG
R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A H

TGAGAAGTCACCAAAACCACAAAATAACCTTACAGCC TTCAGGGAAGAC TACCAGCTCTGCTTTCTACCCCTCTAATTTAACAATGCACCGGAATTCAG
 ACTCTTCAGTGGTTTGGTGTTTTATTGGAATGTCGGAAGTCCCTTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTTGTTACGTGGCCCTTAAGT

linker? -

L R S H Q T T K I T L C P S G K D Y Q L C L S T L . F N N A P E F

CTTGACTTAACC
—————→ 6013
GAACCTGAATTGG

- linker?

L O L Y

Seq ID No 2: Amino Acid Sequence of Hn-unc-53/2 Protein

[illegible][illegible]

FD 302 (Rev. 11-27-70)

1841	1 P S P S A N H S A P S N S T W G T H A S S S S A A S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1842	1 D G L G P O S V S L H T S C K R S I L I I S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1843	1 V P S P M N S S T G L I A S S R O D S L T P P V H S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1844	1 D S D P H L D A N T L P K G L N V Y T F S O L P ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1845	1 T O S E D A K E W L N S H S A G L O O T A A H S P S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1846	1 S G S S V T S P S G T M P N F S O L A S P S T V T ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1847	1 S L S N P T H L N T H S L S N A D G O V O P Y T O S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1848	1 R F N S S H S L D E K S H T M S R S G S P P O G F E ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1849	1 K V H G S S L S L V S S T L S V Y S T P R K A M A Y ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1850	1 R K L R L R E L D A S O E K R V S A L L T O L T A H A ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1851	1 H L V A F R E O S L G R M T I R L U S L T M T A P V A ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1852	1 D S E L N E L R K T I E L L R K V H A A A V A A I H ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1853	1 V I H T P E L N C R G N G T A O S A D L P I P P V H T ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1854	1 S D S V S I N S A T S H S S V G S H I E S D S K L ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1855	1 R K R N W L R S S P R O A P G R K F S P R S A S S ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT
1856	1 ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT ACCOCCT

[illegible]

SEQ. ID NO 3

Nucleic Acid Sequence of
hu-unc-53/1 gene

Page 1

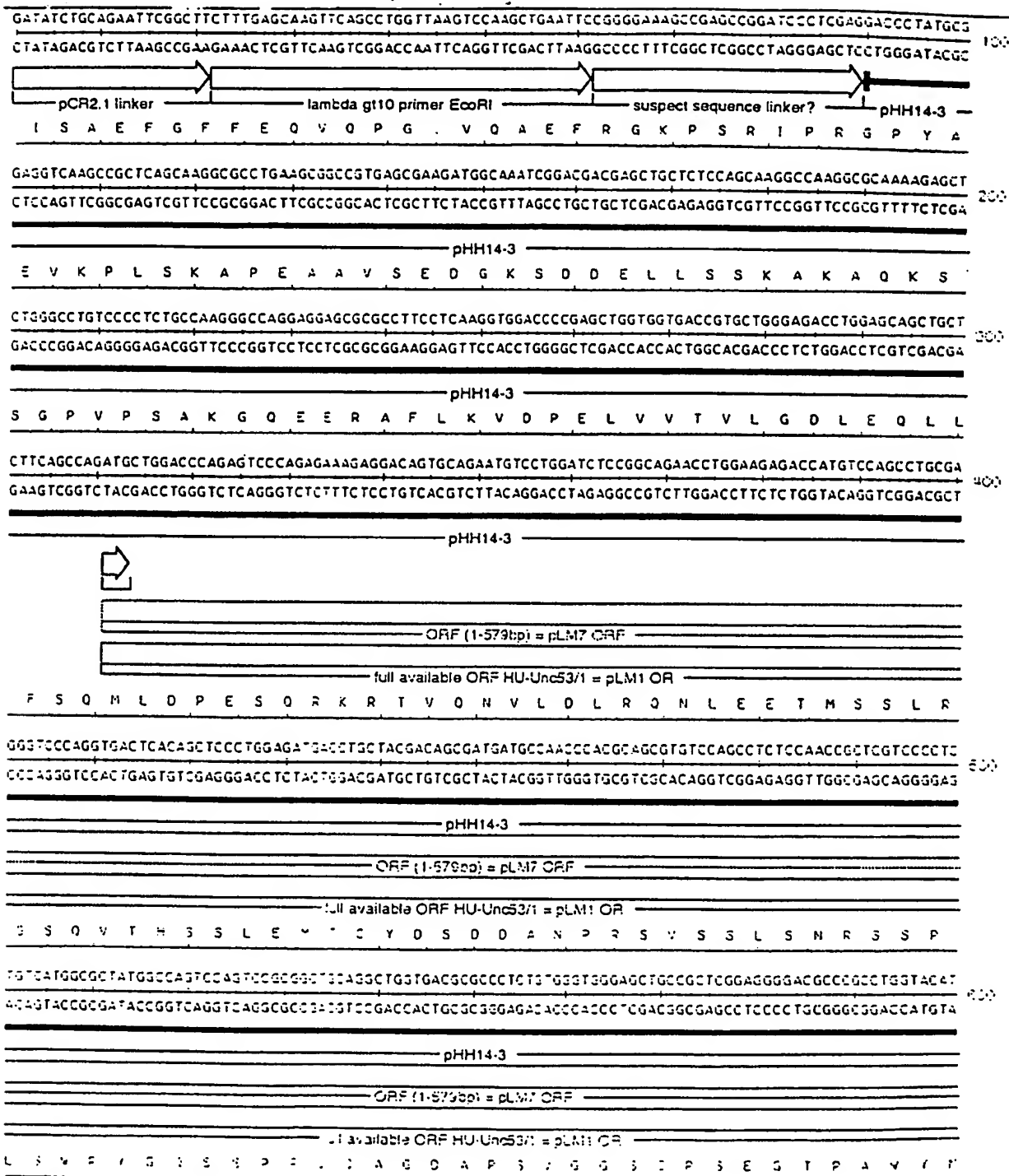


Fig 9

Page 2

Tuesday, 18 November 1997 10:33

fig Hu-Unc53/1 seq (1 > 6013) Site and sequence

GCACGGCGAACGGGCCCACTACTCCACACCATGCCATGCGCAGCCCCAGCAAGCTCAGCCATATC TCCCGCCTGGAGCTGGTCGAATCCC TGGACTCG 700
CGTCCCGCTTGCCCGGGTGATGAGGGTG TGGTACGGGTACGCGCTCGGGGTCTGTCGAGTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC

————— pHH14-3 —————

————— ORF (1-579bp) = pLM7 ORF —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————

H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L D S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACC GGCTGGGATGAAGCA 800
CTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCTCCTACTACTGTAGTGATGGCCGACCC TACTTCTGT

————— pHH14-3 —————

————— ORF (1-579bp) = pLM7 ORF —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————

D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D E S

GCTCCATCAGTAGTGGACTCAGCGATGCCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 900
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTTCATGAGGGTGACG

————— pHH14-3 —————

————— pCB212 —————

————— ORF (1-579bp) = pLM7 ORF —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————

S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P T A

TCTCGCAGGAACCTCAACAATAGTGCCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCC 1000
AAGAGCGTCCCTTGAGTTGTTATCAGGATGCGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTACCCGACTCGACCAATCACTTAGTCTCTCTCTTGGG

————— pHH14-3 —————

————— pCB212 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————

I R R N S T I V L R T C S E K R S L A E S G L S V F S E S E E I A

CCCAAAAAC TGGAGTACGACAGTGGTACGCTGAAAATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTGATCCAGGG 1100
GGATTTTTCGACCTCATGCTGACCATCGGACTTCTACCTTGGACCTGGAAGATTCACCGCTCCCTCGGCGGACTCTGACACTACTAAGTAGGTTCC

————— pHH14-3 —————

————— pCB212 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————

L E I C S C S L V E P G T S K W R P E R P E I C D D S S I

Tuesday, 18 November 1997 10:33

Fig 9

Page 3

lig Hu-Unc53/1 seq (1>6013) Site at Sequence

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCTTGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACTTCCCCATCACTCACAAGG
CACCCTCTTACTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCT 120

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGCCTCAAAGTCGCAGGCAAACCTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGAT
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCTTTTCGATGTCTGTTCCTTCGAACTGACATTTCTTATGACCCGAGGTTGCGAGGAGGAGACTA 130

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

D S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S S D

GCTGGTCGGGACCGCTTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCGCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCTCCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATCTTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTCTTCGGAGGAGGACGGT 140

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTCTATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCCTGTCAGGCCAGTAAATGGGGCCAGGAC
GTCCGTGTCGGTGACAGTACGTTTGACCACCAAGTCCGTGAGAGTCGTCTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGGGTCTCT 150

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M C T G G S A T L S K I O K S S G I P V K P V N S R I T

TAGCTTAGATGTTTCCAACAGTCCAGAGCCAGTATTCCTTGGCTCCTGGAGCCCGTCTTAACAATCCAGTACCGCAGCCCTGCCCCGGCCAGCCAAATCAAGT
ATCGAATCTACAAGAGTGTGACAGTCTCGGTCTTAAGSACCGAGGACCTCGGGCAAGATTGTAGGTGATGCGCTCGGACGGGGCCGGTCCGCTTCAGTTCG 160

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

L D V S H S A E P C F L A F G A R S N I Q Y R S L P R F A S S

Fig 9

Tuesday, 18 November 1997 10:33

Page 7

fig Hu-Unc53/1 seq (1 > 6013) Site a: sequence

TCATGAGCGTGACCGCGGGGGGGTGGTGGCTCGCCCTGTGAGCAGCAGCATTGACCCCACTCTCCACACCAAGCAGGAGAGGCTTACGGCTTCCA
AGATACTCGCACTGGCCGCGCCCGCCACCTGGAGCGGACACTCGTCTGTCTAAGTGGGGTCAGAGGAGTCGTGGTTCTGTCCCTCCGGAATGCGGAAGG

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

rev primer HU53r4

S N S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T P S

GACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGC
CTGACTTCTCGGATGGTTCCATCGGTCACCCGCCCTGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTTCTCTTCCGGTTTCGGTTCCGTCACCC

pHH14-3

pCB212

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

R L K E P T K V A S G R T T P A P V N Q T D R E K E K A K A K A V A

CTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTG
GAACCTGAGTCTGTTGTAGAGGAACCTTCATAACCGAGGGTCTTTCATGAGGGTCTTGGTTCTGTTCTGGTGGGGTGTCTGGTGGTTCGACCGTCTCGAC

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

L D S D N I S L K S : G S P E S T P K N Q A S H P T A T I L A E L

CCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTCTCAACCCACCCCTACTAGCCAATCTTGACAAGGTCAACTCCAAAGCTGTGGATCTACCATCAT
GGTGGTTGGGAGAGTCCCGGTGTCGCTTCTCGAAGCAGTTGGTGGGAGTGTATCGGTTAGAACGTTCAGTTGAGGTGTCTAGACCTAGATGGTAGTA

pHH14-3

pCB210-14

full available ORF HU-Unc53/1 = pLM1 OR

P P T P L R A T A K S F V V P P S L A N L D K V N S N S L D L P S

Tuesday, 18 November 1997 10:33
fig. HU-Unc53/1 seq (1 > 6013) Site. , Sequence

Fig 9

Page 5

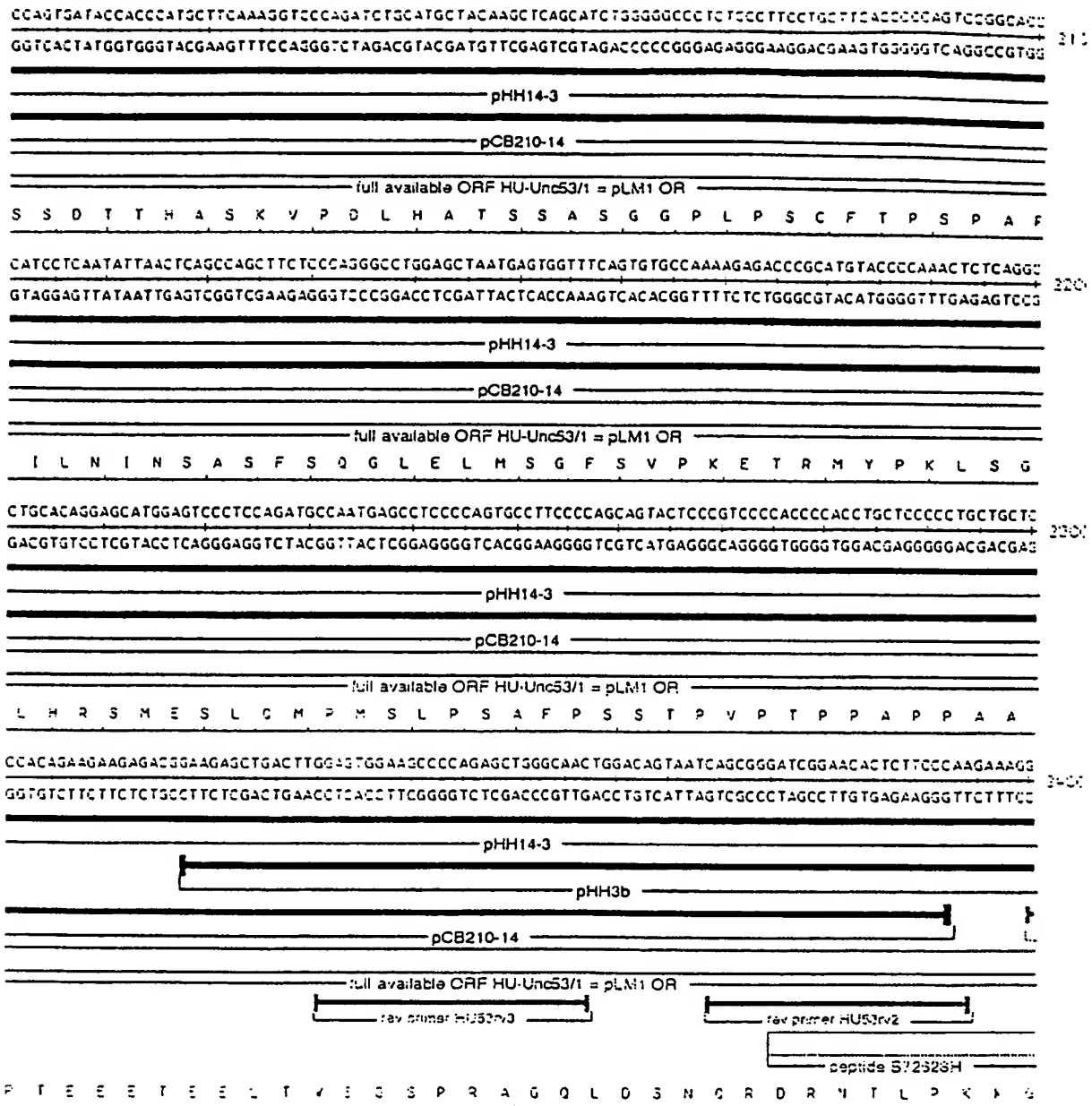


Fig 9

Tuesday, 18 November 1997 10:33

fig. HU-Unc53/1 seq (1 > 6013) Site and sequence

Page 6

GCCTCAGGTACCAGCTTCAGTCCAGGAGGAGACCAAGSAGAGGGGACATTCCCATACCATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT
CGAGTCCATGGTCGAAGTCAGGGTCCCTCTCTGGTTCCTCTCCGCTGTAAAGGGTATGGTAACCCACCGACGGACTTAGGCTACTGGTCAGTCTCGACGGG 350

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E L F

TCTCCCCCTGCACCTTCCCATGTCTCTGAGTGCAGAGGCCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCCAGCCCAAGAACTACCCGCTCCAACA
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGGGTTCTTAGTGGGCGAGGTTGT 360

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P H S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGGGGCTTCGAGCTGTACAGGGCTCCCAAAATGGGGAGCACCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGG
CGTAGGGGTGGGTGCTCCGCCGAAGCTCGACAATGCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTTC 370

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L S G S O M G S T L S L A E R P K G M I R S G

ATCTTCCGAGACCCACGGACGATGTTACGCTTCACTCTCTCTCCCTGGGCTCCAGTGCCTTCTCCACCTACTCTCAGCTGAGGAGAGGATGCAATCT
TAGGAAGGCTCTGGGGTGCCTGTACAAGTGGGAGTCAAGACAGGGACCGGAGGTACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCTCTACGTAGA 380

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

I F R D P T D C V H S S L S L A S S A S S T Y S S A E E R M Q S

GGGCAAAATCCGGAAGCTTCGTAGGGAATGGAAATATCCGAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCGCAATGCTAATCTGGTGGCTGCT
CTGGTTTAGGGCTTCGAAGCATCCCTTGACCTTATACGGTCTCTTTCACCGGTGGAAATGACAGAGTCGAAAGACGTTACGATTAGACCACCGACGAA 390

L2 ORF = pCN351 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F I P K L E E E S E V A T L T S O L S A N A N L V A A

Tuesday, 18 November 1997 10:33

Fig 9

fig HU-Unc53/1 seq (1 > 6013) Site and Sequence

Page:

TTGAGCAGAGCC TGGTGAATATGACATCCCGCCTGCGACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTCGGAGAAACCATAGA
AACTCGTCTCGGACCACTTATAC TGTAGGGCGGACGC TGTGGACCGTCTCTGCGGCTCCTCTTCTGTGACTCGACGACCTAAACGCTCTTTGGTATCT 300

U2 ORF = pC5251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTCTGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTGAGGGAGCCCTTAATGCCCTCAGAAACCACACCCAAAGAAGCTTCGGATCAAGAGACAAAC
GAAAGACTTCTTTTCTTGAGACTCCGGGTCGGTCACTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCCTCTGTTT 310

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I O G A L N A S E T T P K E L R I X R Q H

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCAC TAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGTATGCGAAAAAGAGAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCTTCC TACGACTACGCTTTTCTCTTTTTTCTCAACCC 320

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S ! T S H S S I G S S K D A D A K K K K K S W

TCTATGAGCTTCGAAGTTCTCTCAACAAAGCCTCAGTATAAAAAAGGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCGA
AGTACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCCGGGTTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATGTGGC 330

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K X G P K S A S S Y S O I E E I A T P E

CTCTTCAGCCCCCTCATCCCCCAACTACAGCATGGTTCACAGAGAGTGTTCACCTCCATCAAGTCCCTCCACCTTGTCTTCCCTGGGCACTGATGT
GAGAGTGGGGGAGTAGGGGTTTGATGCTTACCAAGATGCTCTGACCAAGTGGGAGTAGTTCAGGAGGTGGAACAGGAGGACCCCTGACTACAG 340

U2 ORF = pC8251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S P S S P K L Q - S S I E T A S P S I S S T L S S V S T D

Tuesday, 18 November 1997 10:33
fir HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9

Page 8

ACCL .GCCCTGCTACCCAGCCCCACACTAGGC TGTTCATGCAAA TGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGG
TGGCTCCCGGGACGAGTGGGTCCGGGGGTGTGATCCSACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCTTCCATAGCCTCGACSCGAGACTC3
----- U2 ORF = pCB251 ORF -----
----- pHH3b -----
----- full available ORF HU-Unc53/1 = pLM1 OR -----
T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E
TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCACTTGGATCAGCTTCGGGAGACCATGCACAACATGCAGT7
ATACCCTCTTCTTTACTTTCGAATGCTGTAGGCGAACC TCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTGTACGTCAA 3600
----- U2 ORF = pCB251 ORF -----
----- pHH3b -----
peptide B72927H
----- full available ORF HU-Unc53/1 = pLM1 OR -----
----- U3 ORF = pLM5 ORF -----
L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L
GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATT4
CCTCCACCTGGACGACTTTCTGCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAA 3700
----- U2 ORF = pCB251 ORF -----
----- pHH3b -----
----- full available ORF HU-Unc53/1 = pLM1 OR -----
----- U3 ORF = pLM5 ORF -----
E V D L L K A E N D L K V A P G P S S G S T P G Q V P G S S A L
TCTTCCCGACGCCGCTCCCTAGGCC TGGACTTATCCATTCCTTCGGCCCCAGTCTTGACAGACAGACCTGTACCCATGATGGCATCAGTACTTGT3
AGAAGGGGTGCGGCGAGGGATCCGGACGCTGATGGGTAAGGAAGCCGGGTCAGAAGCTCTGTGCTGGACAGTGGGTACC TACCGTAGTCAATGAAC4
----- U2 ORF = pCB251 ORF -----
----- pHH3b -----
----- full available ORF HU-Unc53/1 = pLM1 OR -----
----- U3 ORF = pLM5 ORF -----
S S P R R S L G L A L - S F G P S L A D T C L S P H D G : S T C

Fig. 9

Tuesday, 18 November 1997 10:34

Page 10

fig. 9: Unc53/1 seq (1 > 6013) Site and Sequence

CGTCTGAGGACCGGCGCCTCGTCCCTCGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAA*CGCTTGGCCGAGTACC TGGTGGAGCGCTCTGGG
GGAGGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGTCCCGGTGCCCGTTCGGATGGACTGGTTAGCGAACC GGCTCATGGACCACTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAAC TGTATCTTTCCAACCTAGCCAAC CAGATAGACC
GCACCTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTACAACGTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E V T E S I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAACAGGAATTGGGATGTGCCCCCTGGTGA*TTCTAT TGGATGACCTGAGTGAAGCAGGCTCCATCAGTGA GTTGGTCAATGGGGCCCTCACCTGCAA
GGTGTGTGCTT*AAACCCCTACACGGGGACCACTAAGATAAACC TACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCA GTTACCCCGGGAGTGGAGGT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P E I S I G C V P L L L C D L S E A G S I S E L V Y G A L T C I

Tuesday, 18 November 1997 10:34

Fig 9

Page 11

fig HU-Unc53/1 seq (1>6013) Site Sequence

GTATCAAAATGTCCCTATATTATAGGTACCAACCAATCAGCCCTGTAAAAATGACACCCAAACCATGGCTTGCACCTGAGGCTTCAGGATGTTGACCTTCTCC
CATAGTATTACAGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAACGTGAACCTGAAGTCTTCAAC TGGAGAGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide 872525H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTC
TTGTTGCACCTCGGTCSGTACCGAAGGACCAAGCAATGGACCTCTCTTCGACCATCTCAGTCTGTGCTGTAGTTACGGTGTTCCTTCTCGACGAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGTGGTGGAGTGGGTACCAAGCTGTGGTATCACTCCACACCTTCTTGTAGAACACAGCACCTCAGACTTCTCTCCTGGCCCTTGGCTTCTTCTGT
CCACGAGGTGACCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTCTCTGTGTCTGGAGTCTGAAGGAGTACCGGGGAACGAAGAAAGAAAG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

P V L G A V P K L A I - - P T F - E X H S T S D F L I G P C F F L S

Tuesday, 18 November 1997 10:34
fig. 9 HU-Unc53/1 seq (1>6013) Site an. sequence

Fig. 9

page 12

GTGTCCCATTTGGCATTTGAGGACTTCCGGACCTGGGTTCAATGACCTGTGGAACAACCTATCATCTTCCCTATCTACAGGAAGGAGCCCAAGGATGGGATAAAC
CAGAGGGTAAACCGTAACCTCCTGAAGGCCCTGGACCAAGTAACCTGGACACCTTGTGAGATAGTAAGGGATAGATGTCCCTTCTCGGTTCCCTACCTATTTCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K Q G I I

GTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCCTGCECCATCAGCCCAACAAGACCAATCAAAGCTGTACCACT
CAGGTACCTGTCTTTTCGACGAACCTCCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTGTCTTGGTTASTTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V - G C K A A V E D P V E V V R D T L P W P S A Q Q D Q S K L Y H

TGCCCCCAGCCACCGTGGGCCCTCAGCATTGCCCTCAGCTCCCAGGATAGGACAGTCAAAAGACAGCACCCCAAGTCTCTGGACTCAGATCTCTGAT
AGGGGGTGGGTGGCAGCCCGGAGTGTGTAACGGAGTGGAGGGCTCCTATCTGTGAGTTTCTGTGTGGTGGGTTCAAGAGACCTGAGCTAGGAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

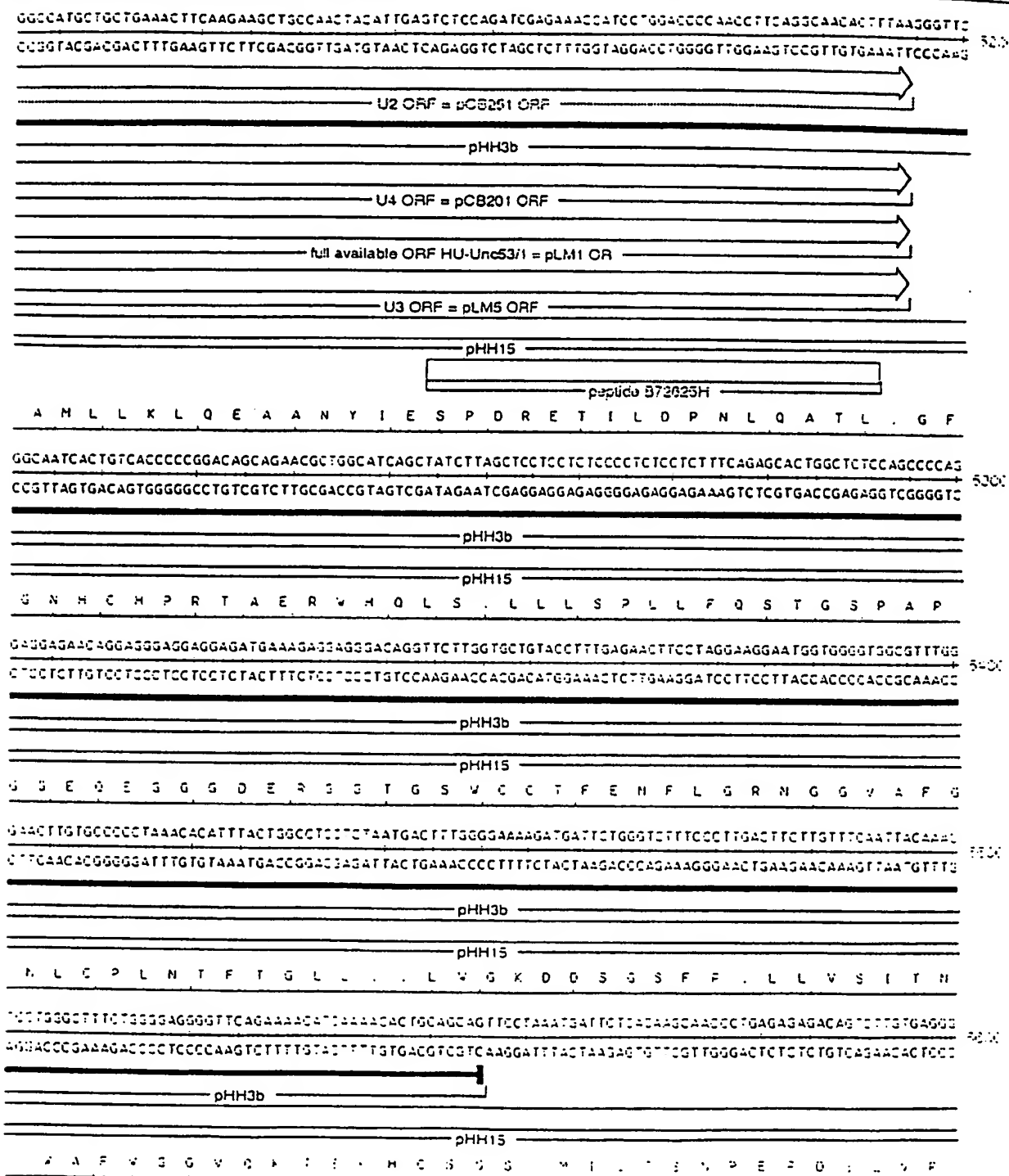
L P P F T V G P F S I A E P P E O R T V K D S T P S S L D S C P L I

Tuesday, 18 November 1997 10:34

fig	tu-Unc53/1 seq (1 > 6013)	Site a	Sequence
-----	---------------------------	--------	----------

Fig. 9

Page 13



Tuesday, 18 November 1997 10:34
fig 1Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9.

Page 17

AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTCTCACAGACCCCTCCCAATTCCATCACCCTGCCAACAACCTCTCCCCAGAGATCTGGCTGGAGC
TCTAGACCCCCCTCCGTCCTTCGAGGAGTCTAAAAGAGTGCTCGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGGTCTCTAGACCGACCTCG 570

PHH15

E I V G R Q E A P Q I F S Q T L P N S I T T A N N S S P R D L A G A
CCAGAAAAAGAAGCATGTGGTTTAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAACAACAAAGCAACAACAAAAACAATGGAA 580
GGTCTTTTCTTCGTACACCAAAATTTTACAAATTTAGTTAGACATTTTCCATTTTACTTTTTGTTTTGTTTCGTTTGTGTTTTGTTTACCTTT
Q K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q V I
AGATGAAGCTGGAGAGAGAGGAACCAAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAAC 590
TCTACTTCGACCTCTCTCTCCTTGGTCAACGGTTCCATCTCTCGACGGGCGAGGACGGGAGACCTACTGTATCCCTGTAGTTGTTCTGCCGACGGTTGG
R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A H
TGAGAAGTCACCAAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTTCTACCTCTAATTTAACAATGCACCGGAATTCAG 600
ACTCTTCAGTGGTTTGGTGTTTTATTGGAATGTCGGAAGTCCCTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC
linker?
L R S H Q T T K I T L Q P S G K D Y Q L C L S T L . F N N A P E F S
CTTGGACTTAACC 6013
GAACCTGAATTGG
linker?
L D L T

[illegible][illegible]

SEQUENCE ID NO 5

[illegible]

118000 2136

SEQUENCE ID NO. 6

Tuesday, 18 November 1997 13:57

(ig 54 pLM1 (1 > 8285) Site and Sequence

Enzymes: 72 of 148 enzymes (Filtered)

Settings: Circular, Certain Sites Only, Standard Genetic Code

Page 1

GTGGCACTTTTGGGGAATGTGGCGGAACCCCTATTTTATTTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATAACCCGTGATAAT
CACCGTGAAAGCCCTTTACACGGCCTTGGGGATAACAATAAAAGATTATGTAAGTTTATACATAGGCGAGTACTGTATTGGGACTATTTA 100
G G T F R G N V R G T P I C L F F . I H S N H Y P L M R Q . P . . M
GTTCAATAATATTGAAAAAGGAAGATGATGATTTCAACATTTCCGTGTCGCCCTTATTCCTTTTTCGGCAATTTGCCCTTCTGTTTGTCTAC
CSAAGTTATTATAACTTTTCTCTCATCTCATAAGTTGTAAGGCAACGCGGAATAAGGAAAAACCGCGTAAACGGAAGGACAAAAACGAGTG 200
L Q . Y . X R K S M S I Q H F R V A L I P F F A A F C L P V F A H
CCGAAACGCTGGTGAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAAGTGGATCTCAACAGCGGTAAAGATCTTGAGAGTT
GGTCTTTGGGACCACTTTTCTTCTACGACTCTAGTCAACCCACGTCCTCACCACATGTAJCTTGACCTAGAGTTGTGCGCCATCTAGGAACCTCAA 300
P E T L V X V K D A E D O L G A R V G Y I E L D L N S G K I L E S
TTGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCGCTATGTGGCGCGGTATTATCCCGTATTGACGCGGGCAAGAGCAACTCGGTGG
AAGCGGGGCTTCTGCAAAAGGTTACTACTCGTGAAATTTCAAGACGATACACCGGCCATATAGGGCATAACTGCGGGCCGCTTCGTTGAGCCAGC 400
F R P E E R F P M H S T F K V L L C G A V L S R I D A G Q E Q L G R
CCGCATACACTATTCTCAGATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCC
GGGTATGTGATAAGAGTCTTACTGAACCAACTCATGAGTGGTCACTGCTTTTCGTAGAATGCTACCGTACTGTCTTCTTAATACGTACGACGG 500
R I H Y S Q N O L V E Y S P V T E K H L T D G H T V R E L C S A A
ATAACCATGAGTGATAACACTCGCGCCCACTTACTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTCACAACATGGGGATCATGTAA
TATTGGTACTCATATTGTGACGCGCGTGAATGAAGACTGTGCTAGCCTCTCGGCTTCCGATGCGGAAAAAGCTGTGTGACCCCTAGTACATT 600
I T H S O N T A A N L L L T T I G G P K E L T A F L H N M G D H V
CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCATAACCAACGACGAGCGTGACACCGATGCCGTGAGCAATGGCAACAGTTGCGCAAACT
GAGCGGAACTAGCAACCTTGGCTCGACTTACTTCGGTATGCTTGTGCTGCGCACTGTGGTGTACGGACATCGTTACCGTTGTGCAACGCGTTTGA 700
T R L D R V E P E L N E A I P N D E R D T T N P V A M A T T L R K L
ATTAACTGGCGAATCTTACTCTAGCTTCCCGCAACAATTAATAGACTGGATGGAGCGGATAAAGTTGACGAGCACCTTCTGCGCTCGGCCCTTCGG
TAATTGACCGCTTGATGAATGAGATCGAAGGGCGTGTGTAATTAATGACCTACCTCCGCTATTTCACAGCTCTGGTGAAGACGCGAGCGGGAAGCG 800
L T G E L L T L A S R O O L I D V M E A D K V A G P L L R S A L P
GCTGCTGCTTTTATGCTGATAAATCTGGAGCGGCTGAGCGTGGGTCTCGCGGTATCATTCGACACTGGGGCCAGATGGTAAGCCCTCCGCTATCGTAG
TAAAGGACCAATAACGACTATTATAGACTCGGCACTCGGCACTCGGCACTAGAGGCACTAGTAACCTGCTGACCCCGGCTTACCATTCGGGAGGCAATAGCATC 900
A G V F I A D K S G A G E R G S R G I I A A L G P O G K P S R I V
TAAATACAGACGGGAGTCAGGCAACTATGGATGAACSAATAAGACAGATCGCTGAGATAGGTCCTCACTGATTAAACATTTGTAAGTTCAGACCA
AATAGATGCTGCGCCCTCAGTCCGTTGATACCTACTTGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
V I Y T T G S O A T M D E R N R O I A E I G A S L I K H V . L S D O
AGTTTACTCATATATCTTATAGTTGATTTAAAACCTTCATTTTAAATTAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCT
TCAATGAGATATATGAATCTAACTAAATTTTGAATTAAGAAATTAATTTTCTTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGGTTTAGGGA 1100
V Y S Y I L . I D L K L H F . F K R I . V K I L F O N L M T K I P
TAAGTGAATTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTCGCGCGTAATCTGCTGCTTGCAAA
ATTGCACTCAAAAGCAAGGTGACTGCACTGCTGGGGCACTTTTCTAGTTTCTAGAAGAACTCTAGGAAAAAGACGCGCATTAGACGATCAACGTTT 1200
R E F S F H . A S D P V E K I K G S S . O P F F L R V I C C L Q
CAAAAAACCACTGCTACACGCGGTGGTTTGTGCT 1300
GCTTTTGGTGGGATGGTGGGCAACCAACAGGCTTATGCT
T A K P P L P A V V C L P O Q E L P T L F P K V T G F S R A C I P N
TATGCTGCTTCTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCTTAGTGTAGGCT 1400
AATAGAGGAT
V L L V . P L G M H F N S V A P P T L A L L I L L P A A

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 2

GCAGTGGCGATAAGTCGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGCTGAACGGGGGTTCTGTCACACAGC 1500
CGGTACCCGCTATTACGACACAGAAATGGCCCAACCTGAGTCTGCTATCAATGGCCATTTCGCGGTGCGCAGCCGACITGCCCCCAAGCACGTGTGTG
A S G D K S C L T G L D S R R . L P D K A Q R S G . T G G S C T Q
CCAGCTTGGAGCGAACGACCTACACCGAAGTACGATACCTACAGCCTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGCGGACAGGTATCC 1800
GTCGGAACCTCGCTTGTGATGTGGCTTGACTCTATGATGTGCGACTCGGATACCTTTTCGCGGTGCGAAGGGCTTCCCTCTTTCCGCTTGTCCATAGG
P S L E R T T Y T E L R Y L O R E L . E S A T L P E G R K A D R Y P
GGTAAGCGCGAGGGTCGGAACAGGAGAGCGCACGAGGAGCTTCCAGGGGGAACGCTGGTATCTTTATAGTCTGTGCGGTTCGCGACCTCTGACTT 1700
CCATTCGCGCTCCGACGCTTGTCTCTGCGGTGCTCCCTCGAAGGTCCCTTTTCGCGACCATAGAAATATCAGGACAGCCAAAGCGGTGGAGACTGAA
V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L . L
GAGCTCGATTTTGTGATGTCTGAGGGGGGCGAGCTATGGAAGACCGCAGCAAGCGGCCCTTTTACGGTTCCTGGCTTTTGTGCGCTTTTG 1800
CTGCGAGCTAAAGAACTACGAGCAGTCCCTCGCTCGGATACCTTTTTCGCGGTGCTGCGCGGAAAGATGCCAAGGACCGGAAAGACCGGAAAGC
E R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F
CTCAGATGTTCTTTCGCGCTATCCCTGATTCGTGGATAACCGTATTACCGCTTTTGTGAGCTGATACCGCTGCGCGCAGCGAAGCAGCGAGCG 1900
GAGTGTACAAAGAGGACCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACCTGACTATGGCGAGCGCGCTCGCTTGTGCTGCGT
A H M F F P A L S P D S V D N R I T A F E . A D T A R R S R T T E R
CAGCGAGTCAGTACGAGGAGGAGCGGAAGACCGCCCAATACGCAACCGCCTCTCCCGCGGTGGCGGATTCAATATGACAGCTGGCAGCAGGTTT 2000
GTGCTCAGTCACTCGCTCTTTCGCTTCTCGCGGGTATGCGTTTGGCGGAGAGGGGCGCGCAACCGGTAAAGTAATACGTCGACCGTGTGTCGCAAA
S E S V S E E A E E R P I R K P P L P A R V P I H . C S V H D R F
CCGCTGGAAGCGGCGAGTACGCGCAACGCAATTAATGTAGTACCTCAATTAGCGACCCAGGCTTTACACTTTATGCTTCCGCTCGTATGT 2100
GAGCTGACCTTTTCGCGCTGCTCGGTTGCGTTAATACACTCAATCGAGTGAGTAATCGTGGGTTCGAAATGTGAAATACGAGGCGGAGCATACA
P D V K A G S E R N A I N V S . L T H . A P Q A L H F H L P A R M
TGTGTGAATGTGAGCGGATAACAATTCACAGGAAACAGCTATGACCATGATTACGCAAGCGCGCAATTAACCTCACTAAAGGGAACAAAAGCT 2200
ACACACCTTAACACTCGCTATGTTAAAGTGTCTCTTGTGGA*ACTGGTACFAATGCGGTTCGCGGTTAATGGGAGTGATTTCCCTTGTGTTTGA
L C G I V S G . Q F H T G N S Y D H O Y A K R A I N P H . R E Q K L
GGGTACCGGGCCCCCTCGAGGTGACGCTATCGATAGCTTATCGAATTCCTGACGCCCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCAG 2300
CCATATGGCCCGGGGGAGCTCCAGCTGCTATAGCTATTCTCACTA*AGCTTAAGGAGCTCGGGGACGAGAAGTCCGCTACGACCTGGGCTCAGGGTC
insert pLM1
ORF pLM1
G T G P P L E V D G I D < L D ! E F L Q P L L F S Q N L D P E S Q
AGAAAGAGGACAGTGCAGAAATGCTCGGATCTCGGCGAGAACTTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCGAGGTGACTCACAGCTCCCTGGAGA 2400
TCTTTCTCTGTCACGCTTACAGGACCTAGAGGCGCTT*GACCTTCTCTGGTACAGGTGCGACGCTCCAGGGTCCACTGAGTGTGAGGGGACCTCT
insert pLM1
ORF pLM1
R K R T Y Q N V L D L R D N L E E T M S S L R G S O Y T H S S L E
TGACCTGCTACGACAGCGATGATGCCAAGCCACGACGCTT*CCAGCTCTCAACCGCTGCTCCCTCTGTCATGGCGCTATGGCCAGTCCAGTCCGCG 2500
AT*GGACGATGCTGCTGCTACGCTTACGGTTGGGTGGCTGCT*ACAGCT*GGAGAGGTTGGCGAGCAGGGGAGACAGTACCGGCTACCGTCAAGTCAAGGCG
insert pLM1
ORF pLM1
M T C Y D S D O A N P R S V S S L S N R S S P L S V R Y G O S S P R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 6

```
CC TGGGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTGCGAGAAACCATAGACTTCTGAAGAAAAGAACCTGAGGCCCGAG 5000
GGACGCTGTGGACCGCTCTGCGCGGCTCTCTTCTCTGTGACTGAGACCAACGCTCTTTGGTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTC
-----insert pLM1-----
-----ORF pLM1-----
L R H L A E T A E E K D T E L L D L R E T I D F L K K K N S E A Q
GCAGTCATTAGGGAGCCCTTAATGCCTCAGAAACCAACCTAAAGAACTTCGGATCAAGAGACAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCA 5100
CGTCAGTAAGTCCCTCGGGAATACGGAGTCTTTGGTGTGGGTTCTTGAAGCCTAGTCTCTGTTTGGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGT
-----insert pLM1-----
-----ORF pLM1-----
A V I Q G A L N A S E T T P K E L R I K R Q N S S D S I S S L N S
TCAC TAGCCATTCAGCATCGGCAGCAGCAAGGATGCTGA*GGGAAAAGAGAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCTTCAACAAAGC 5200
AGTGATCGGTAAAGTCGTAGCCGTCGTGCTTCTACGACTACGCTTTTCTTCTTTTCTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCG
-----insert pLM1-----
-----ORF pLM1-----
I T S H S S I G S S K D A D A K K K K K S V V Y E L R S S F N K A
GTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCAGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAACTACAG 5300
CAAGTCATATTTTTCCTCCGGGTTTCAGTCAAGGAGTATGAGCTATATCTCTCTTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGGTTGATGTC
-----insert pLM1-----
-----ORF pLM1-----
F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L Q
CATGGTTCACAGAGAGCTGCTTACCCCTCATCAAGTCTCTACCTTGTCTCCGTCGGGCTGATGTACCGAGGGCCCTGCTACCCAGCCCCCACA 5400
GTACCAAGGTGTCTCTGACCAAGTGGGAGGTAGTTTCAGGAGGTTGAGCAGGAGGACACCGTGACTACAGTGGCTCCCGGACGAGTGGGTGGGGGGTGT
-----insert pLM1-----
-----ORF pLM1-----
H G S T E T A S P S : K S S * L S S V G T D V T E G P A H P A P H
TTAGGCTGTTCATGCAATAGGAGGAGGACCTAGAGAAAGAGGATATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACAT 5500
GATTCGACAAGGTACGTTTACTGCTCTCTCTGAGTCTCTTCTCTCTCATAGCTTCGACGCGAGACTCGATACCTCTCTCTTACTTCGAATGCTGTGA
-----insert pLM1-----
-----ORF pLM1-----
I R L F H A N E E E P E K Y E V S E L R S E L V E K E M K L T D I
TTCTTGGAGGGCTCAACTCTGCGCAACCACTGATGAGCTTTCTGAGACCATGCACAACATGCAAGTGGAGGTGGACC TGTGAAGCAGAGAATGAC 5600
TGGCAACTTCGGGAGTTGAGACGGGTGGTGAAGTATGAGTGAAGTCTCTGAGTACGTGTTGACGTCAACCTCCACCTGGACGACTTTCGTCCTTACTG
-----insert pLM1-----
-----ORF pLM1-----
P L E A L N S A M D L D G L E T H H N H O L E V D L L K A E N D
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT 5700
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT
-----insert pLM1-----
-----ORF pLM1-----
P L E A L N S A M D L D G L E T H H N H O L E V D L L K A E N D
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT 5800
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT
-----insert pLM1-----
-----ORF pLM1-----
P L E A L N S A M D L D G L E T H H N H O L E V D L L K A E N D
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT 5900
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT
-----insert pLM1-----
-----ORF pLM1-----
P L E A L N S A M D L D G L E T H H N H O L E V D L L K A E N D
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT 6000
TATGTAAGGTAGCCCTAGGCCCTCATCAGGCTCCCACTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGTCTCTGAGT
```


Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

ATGAGGGACAGGTTCTTGGTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTTGCCCCCTAAACACATTACTGGC
TCTCCCTGTCCAAGAACACGACATGGAAACTCTTGAAGGATCTTCTTACCACCCACCGCAACCCCTTGAACACGGGGATTGTGTAATGACCG 7400

insert pLM1

G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G

CTCTCTAATGACTTTGGGAAAAGATGATTCTGGGTCTTCCCTTGACTTCTTGTTCATTACAACTCCTGGGCTTCTGGGGAGGGGTTGAGAAA 7500

GAGGAGATTACTGAAACCCCTTTCTACTAAGACCCATAAGGGAAGTGAAGAACAAGTTAATGTTTGGAGCCCGAAAGACCCCTCCCAAGTCTTTT

insert pLM1

L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K

CATCAAACTGTCAGCAGTTCCECGGAATTCAGCTTTCACCTAACAGGCTGAACCTTGCTCAAAAGAAGCCGAATTCAGCACACTGGCGCCGTTACT 7600

GTAGTTTGTGACGTGCTCAAGGGGCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGTTTCTTCGGCTTAAGGTCGTGTGACCGCCGCAATGA

insert pLM1

T S K H C S S S P E F S L D L Y R L N L L K R S R I P A H V R P L L

AGTTCTAGAGCGGCGCCACCGCGGTGGAGCTCCAATTGCGCTATAGTGAGTCGTATTACGCGCGCTCACTGGCGTCTGTTTACAACTGTCGACTGG 7700

TCAAGATCTCGCGCGGTGGCGCCACCTCGAGGTTAAGCGGATATCACTCAGCATAATGCGCGGAGTGACCGGCAGCAAAATGTTGACGACTGACC

V L E R P P P R V S S N S P Y S E S Y Y A R S L A V V L O R R D V

GAAAACCTGGCGTTACCAACTTAATCGCCTTGACGACATCCCTTTTCGCCAGCTGGCGTAATAGCGAAGAGCGCCGACCGATCGCCCTTCCCAAC 7800

CTTTGGGACCGCAATGGTGTGAATFAGCGGAACGCTCTAGTGGGGAAGCGGTCGACCGCATTATCGCTTCTCGGGCTGGCTAGCGGGAAGGGTTG

E N P G V T Q L N R L A A H P P F A S V R N S E E A R T D R P S Q

AGTTGCGCAGCTGAATGGCGAATGGGACGCGCCCTGTAGCAGCGATTAGCGCGCGGGGTGGTGGTGTACGCGCAGCTGACCGCTACACTTGCCAG 7900

TCAACGCGTCGGACTTACCGCTTACCTTGCGCGGGACATCGCCCGTAATTCGCGCGCCACACCAATGCGCTCGCACTGGCGATGTGAACGGTC

Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S

CGCCCTAGCGCCGCTCTCTTCGCTTCTTCCCTTCTCTTCTGCGCAGTTGCGCGGCTTCCCGTCAAGCTCTAAATCGGGGCTCCCTTTAGGGTTC 8000

CTGGGATCGCGCGGAGGAAAGCGAAAGAGGGAAGGAAAGAGCGGTGCAAGCGGCGGAAAGGGGAGTTGAGATTAGCCCCGAGGGAATCCCAAG

A L A P A P F A F F P S F L A T F A G F P R O A L N R G L P L G F

CTATTTAGTGCTTACGGCACCTCGACCCCAAAAACCTTATTAAGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGTTTTTCGCCCTTTGA 8100

CTTAATCAGGAAATGCGGTGGAGCTGGGGTTTTTGAATTAATCCCACTACCAAGTGCATACCCGGTAGCGGGACTATCTGCCAAAAGCGGGAATCT

R F S A L R N L D P K K L D . G D G S R S G P S P . . T V F R P L

CTTTGAGTCCACGTTCTTTAATAGTGGACTCTTGTCTCAAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTTGATTATAAGGGATTTTGGC 8200

GCAACCTCAGGTGCAAGAAATATCACTGAGAACAAGTTTGAACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAATAAATATTCCTTAAACCG

I L E S T F F N S G L L F D T G T T L N P I S V Y S F O L . G I L P

CTTTTCGGCTTATGGTTAAAAAATGAGCTGATTTAACTAAAAATTAACCGGAATTTTAAACAAATATTAACGCTTACAATTTAG 3295

CTTAAGCGGATAACCAATTTTCTACTGCACTAAATTTCTTAAATTCGCTTAAATTTGTTTATAATTCGCAATGTTAAATC

: S A Y V L K H E L I . : < F N A N F N K I L T L T I .

SEQUENCE ID PCT

Page 1

Tuesday, 18 November 1997 11:48

fig 34 pLM4 (1 > 10070) Site and Sequence

Enzymes : 100 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

100
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCC
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGC
PCMV
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
200
CCCCAACGACCCCGCCCATTCAGCGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
PCMV
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G V F T V
300
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCATATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCAT
PCMV
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
400
CATGACCTTATGGGACTTTCCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACACGCGCAAAACCGTCATGTAGTTACCCGACCT
PCMV
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V
500
TAGCGGTTTGACTCAGGGGATTTCCAAGTCTCCACCCCATTCAGCGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCACAAATGTCGTA
ATCGCAAACTGAGTGCCCTAAAGGTTTCAAGGTTGGGGTAACGTCAGTTACCTCAAAACAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT
PCMV
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .
600
ACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGTAACGCGTTTACCCGCCATCCGACATGCCACCTCCAGATATATTCGTCTCGACCAAACTCACTTGGCAGTCTAGGCGATCGCGAT
PCMV
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
700
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGGG
GGCAGCGGTGGTACCACTCGTTCCCGCTCCCTGCAAAAGTGGCCCCACACGGGTAGGACCAAGCTCGACCTGCCGCTGCATTTCGCGGTGTTCAAGTCGC
EGFP
P V A T M V S K G E E L F T G V V P I L V E L D G O V N G H K F S
800
TGTCCGGGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTGATCTGCACACCGGCAAGCTGCCCGTGCCCTGGCCCAACCTCTGTGAC
AGAGGCCGCTCCCGCTCCCGCTACGGTGGATGCGCTTCTGAC TGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGACCGGGACCGGGTGGGAGCACTG
EGFP
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
900
GACCGTGACCTACGGCGTGCGAGTGCCTCAGCCGCTACCCCGACCATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGAG
GTGGGACTGGATGCCCGACCTCAGCAAGTGGGCAATGGGCTGGGTACTTTCGTCGTCTGAAGAAGTTACGGCGGTACGGGCTTCGATGCGAGTCTCTC
EGFP
L T T G V G C F S R . S D H N K O H D F F K S A M P E G Y V Q E

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 100

EGFP

R T I F F K O D G N Y K T R A E V K F E G D T L V H R I E L K G I

ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAAGCGCATCAA
TGAAGTTCTCTGCGGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCGTCTCTTGGCGTAGTT 110

EGFP

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I A

GGTGAATTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCCTGCCGTGCGACGTCGAGCGGCTGGTGTGGTCTGTTGTGGGGTAGCCGCTGCCGGGGCACGACGAC 120

EGFP

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCATCACATGGTCTGCTGGAGTTCGTGACCGCGCCGGGA
GGGCTGTGGTGTGGACTCGTGGGTGAGGCGGACTCGTTTCTGGGGTGTCTCTTCGCGCTAGTGTACAGGACGACCTCAAGCACTGGCGGGCGCCCT 130

EGFP

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGACTCAGATCTCGAGCTCAAGCTTCGAATCTGCAGTCGATAAGCTTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTTACGGCTTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTGCAACTATAGCTTAAGGACGTCGG 140

EGFP

I T L G M D E L Y K S G L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAAGAGAGGACAGTGCAGAAATGTCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCCTGTACAGCTTACAGGACCTAGAGGCGGCTCTGGACCTTCTCTGGTACAGGTCT 150

insert pLM1

ORF pLM1

L L F S Q M L D P E S Q R K R T V Q N V L D L R O N L E E T M S S

CTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGACGCTGTCCAGCCCTCTCAACCGCTCG
GACGCTCCAGGGTCCACTGAGTGTGAGGGACCTCTATGGACGATGCTGTCGCTACGCGGTGGGTGCGTCCGACAGGTCGGAGAGGTTGGCGAGCA 160

insert pLM1

ORF pLM1

L R G S Q V T H S S L E M T C Y D S D D A N P R S V S S L S N R S

CCCCCTGTGTCATGGCGCTATGGCCAGTCCAGTCCGCGCTGACAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCT
GGGAGACAGTACCGGATACCGGTACGGTCAAGGCTGAGGCTGACGCGGAGACACCCACCTTCGACGCGAGGCTCCCTCTCGGGCGGAC 170

insert pLM1

ORF pLM1

L L S R R I G Q S S R L G A G D A P S V G G S C R S E G T P A V

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 3

GACATGCACGGCGAACGGGCCCACTACTCCACACCATGCCCATGGCGACCCCAAGCTCAGCCATATCTCCCGCCGGGAGCTGGTGGAAATCCCTG
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGCGTCGGGGTCGTTCGAGTCGGTATAGAGGGCGGACCTCGACCAAGCTTAGGGAC

insert pLM1

ORF pLM1

Y M H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L

GACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG
CTGAGCCCTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCACTGGAGTACCCGTTCTGGTACTGCCCTCTACTACTGTAGTGATGGCCGACCTAC

insert pLM1

ORF pLM1

D S D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D

AAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCCTCAGACAATCTCAGTTCAGAAGAATCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCC
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTTCATGAGG

insert pLM1

ORF pLM1

E S S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P

CACTGCTTCTCGCAGGAACCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTAGTGAATCAGAGGAG
GTGACGAAGAGCGTCTTGAGTTGTTATCACGATGCTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAACTCACTTAGTCTCCTC

insert pLM1

ORF pLM1

T A S R R N S T I V L R T D S E K R S L A E S G L S W F S E S E E

AAAGCCCTAAAAAAGTGGAGTACGACAGTGGTAGCCTSAAGATGGAACCTTGGACTTCTAAGTGGCGGAGGGAGCGGCCCTGAGAGCTGTGATGATCA
TTTCGGGATTTTTTGACCTCATGCTGTACCAACGACTTCTACCTTGGACCTGAAGATTACCGGCTCCCTCGCCGGGACTCTCGACACTACTAAGTA

insert pLM1

ORF pLM1

K A P K K L E Y C S G S L K M E P G T S K V R R E R P E S C D D S

GCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCGACCTGGTTCCCTGAAGAAGGCGCAAGACCCCACTGTGGCTGTAACCTCCCGCATCACTCA
GGTCCCACTCTTGACTTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGACTTCTCCCGTTCTGGGGTGGACACCGACATTGAAGGGGTAGTGAAT

insert pLM1

ORF pLM1

G G E L K K P I S L S P G S L K K G K T P P V A V T S P I T H

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page L

CACAGCCCAGAGTGGCC TCAAAGTCGCAGGCAAACTGAGGGCAAAGCTACAGACAAGGTAAGCTTGCAGTGAGGAATACTGGGCTCCAAACGCTCCTCC
GTGTCGGGTC TCACGGGAGTTTACAGCTCCGTTTGGACTCCCGTTTCGATGCTGTGTTCCCATTCGAACGTCACCTTCATGACCCGAGGTTCGAGGAGG

2400

insert pLM1

ORF pLM1

T A O S A L K V A G K P E G K A T D K G K L A V K N T G L O R S S

TCTGATGCTGGTCGGGACCGCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTC
AGACTACGACCAGCCCTGGCGGACTCAC TACGATTCTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAG

2500

insert pLM1

ORF pLM1

S D A G R O R L S D A K K P P S G I A R P S T S G S F G Y K K P P

CTGCCACAGGCACAGCCACTGTCTATGCAAAC TGGTGGTTACAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCG
GACGGTGTCGGTGTGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCCTCAGGAGTCCGTAGGGACAGTTTCGGTCAATTACCCGC

2600

insert pLM1

ORF pLM1

P A T G T A T V M O T G G S A T L S K I O K S S G I P V K P V N G R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAAG
GTCTGATCGAATCTACAAAGGTTGTCGGCTCTGGTCC TAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGCGGGTGGTTTC

2700

insert pLM1

ORF pLM1

K T S L D V S N S A E P G F L A P G A R S N I O Y R S L P R P A I

TCAAGTTCTATGAGCGTGACCGGCGGGGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCCTCAGCACCAAGCAGGAGGCGCTTACGG
AGTTCAAGATACTCGCACTGGCCGCGCCGCCACCTGGAGCGGGACACTCGTCGTCGTAAC TGGGTCAGAGGAGTCGTGGTTCGTCCTCCGGGATGCG

2800

insert pLM1

ORF pLM1

S S S M S V T G G R G G P R P V S S S I D P S L L S T K O G G L T

CTTCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCACTCCAGCCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGG
GAAGGTCGACTTCTTCGGATGTTTCCATCGGTCACCCGCTGGTGAGGTCGGGACAGTTAGTCTGTC TAGCCCTTTCTCTTCCGGTTTCGGTCCG

2900

insert pLM1

ORF pLM1

F S R L K E P T K V A S G R T T P A P V H O T C R E X E K A A A I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

AGTGGCCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAGCAAGCCACCCCAAGCCACCAAGCTGGCA
TCACCGGAACCTGAGTCTGTTGTAGAGGAACCTCTCATAACCGAGGGGTCCTCATGAGGGTTCCTTGGTTCGTTCGGTGGGGTCTCGGTGGTTCGACCGT 300

— insert pLM1 —

— ORF pLM1 —

V A L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A

GAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACACCCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTAC
CTCGACGGTGGTTGGGGAGAGTCCCGGTGTCGCTTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCCAGTTGAGGTGTGCAGACCTAGATG 310

— insert pLM1 —

— ORF pLM1 —

E L P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L

CATCATCCAGTGATACCACCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCAAGTCC
GTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTTCGAGTCGTAGACCCCGGGAGAGGGAAGGACGAAGTGGGGTTCAG 320

— insert pLM1 —

— ORF pLM1 —

P S S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P

GGCACCCATCCTCAATATTAACCTCAGCCAGCTTCTCCAGGGCCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACTC
CCGTGGGTAGGAGTTATAATTGAGTCGGTCTGAAGAGGGTCCCGGACCTCGATTACTACCAAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGA 330

— insert pLM1 —

— ORF pLM1 —

A P I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L

TCAGGCCCTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCTCCCAAGTGGCTTCCCAAGCAGTACTCCCGTCCCAACCCACCTGCTCCCGCTG
AGTCCGGACGTGTCCTCTACCTCAGGGAGGTCTACGGTTACTCGGAGGGGTACGGAAGGGGTCTGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGAC 340

— insert pLM1 —

— ORF pLM1 —

S G L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P

CTGCTCCCAAGAGAAGAGAGCGGAAGAGCTGACTTGGAGTGGAGGCCAGAGCTGGGCAACTGGACAGTAACTACGGGGATCGGAACACTCTTCCCAA
GACGAGGGGTGCTCTCTCTGCTCTCTGACTTAACCTCAGCTTCGGGGTCTCGACCCGTTGACCTGTCAATTAGTCGCTCTGTSAGAAGGGTT 350

— insert pLM1 —

— ORF pLM1 —

A A P T E E E T E E L T V S G S P R A G Q L D S N Q R D R N I L F I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

GAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATCCCATACCATGGTGGGCTGCCATCCGATGACCAGTCAGAG
CTTCCCGAGTCCATGGTCGAAGTCAGGGTCTCTCTGGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTC 360

insert pLM1

ORF pLM1

K G L R Y Q L O S Q E E T K E R R H S H T I G G L P E S O O Q S E

CTGCCCTTCTCCCCCTGCACCTCCCATGTCTCTGAGTGCAAGGGCCAACCTACCAACATAGTGAGTCCCCTGCGGCCACCCAGCCAAGAATCACCCGCT
GACGGAAAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTTGAATGGTGTATCACTCAGGGTGACGCCGGTGGTGGGTTCTTAGTGGGGCA 370

insert pLM1

ORF pLM1

L P S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R

CCAACAGCATCCCCACCCACGAGGCGGCCCTTCGAGCTGTACAGCGGCTCCCAATGGGGAGCACCTGTCCCTGGCCGAGAGACCAAGGGAATGATTCE
GGTTGTCTGAGGGGTGGGTGCTCCGCCGAAGCTCGACATGTGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGC 380

insert pLM1

ORF pLM1

S N S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R

GTCAGGATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGCTGTCCCTGGCCCTCCAGTGCCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATG
CAGTCTTAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTACGACAGGGACCGGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTAC 390

insert pLM1

ORF pLM1

S G S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M

CAATCTGAGCAAAATCCGGAAGCTTCGTAGGGAACGGGAATCATCCAGGAAAAAGTGGCCACCTTGACGTCCTCAGCTTCTTGCCAATGCTAATCTGGTGG
GTTAGACTCGTTTAGGCTTCGAAGCATCCCTTGACCTTAGTAGGGTCTCTTTTCACCGGTGGAACGTCAGAGTCGAAAGACGGTTACGATTAGACCAC 400

insert pLM1

ORF pLM1

D S E Q I R K L R R E L E S S O E K V A T L T S Q L S A N A N L V

CTCCTTTTGAGCAGAGCCTTGTGAATATGACATCCCGCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAAC
GACGAAAACCTCGTCTCGGACCACTTATACTGTAGGGCGSACGCTGTGGACCGTCTCTGCCGGCTCCTCTCTCTGTGAETCGACGACCTAAACGCTCTTG 410

insert pLM1

ORF pLM1

A A F E Q S L V N M T S R L R H L A E T A E E K O T E L L O L R E I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

```
CA TAGAC TTTCTGAAGAAAAGAACTCTGAGGEC CAGGCAGTCATT CAGGGAGCCCTTAATGCC TCAGAAACCACACCCAAAGAACTTCGGATCAAGAGA
G TATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCCTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGGGGTTTCTTGAAGCCTAGTTCTCT 220
-----insert pLM1-----
-----ORF pLM1-----
I D F L K K K N S E A Q A V I O G A L N A S E T T P K E L R I K R
CAAAACTCCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCCGAAAAAGAGAAAAAAGA
G TTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTCTCTTTTTTCTCT 230
-----insert pLM1-----
-----ORF pLM1-----
C N S S O S I S S L N S I T S H S S I G S S K D A D A K K K K K L
GTTGGGTCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCAAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTAC
CAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGAAGTCATATTTTTCCTCCGGGTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATG 240
-----insert pLM1-----
-----ORF pLM1-----
S V V Y E L R S S F N K A F S I K K G P K S A S S Y S O I E E I A T
ACCCGACTCTTCAGCCCCCTCATCCCCAACTACAGCATGGTTCCACAGAGACTGCTTCACCCCTCATCAAGTCTCCACCTTGCTCTCGTGGGCACT
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGTCCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTA 250
-----insert pLM1-----
-----ORF pLM1-----
P O S S A P S S P K L C H G S T E T A S P S I K S S T L S S V G T
GATGTCACCGAGGGCCCTGCTCAGCCAGCCCCACACTAGGCTGTTCCATGCAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCT
CTACASTGGCTCCCGGACGAGTGGGTGCGGGGGTGTCATCCGACAAGGTACGTTTACTCTCTCTCTCGGTCCTTCTTCTCTCATAGCCTCGACGCGA 260
-----insert pLM1-----
-----ORF pLM1-----
D V T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R
CTGAGCTATGGGAGAAGGAATGAAGCTTACAGACATCCCTTGGAGGGCCCTCAACTCTGCCCACTGGATCAGCTTCGGGAGACCATGCACAAACAT
GACTCGATACCCCTTCTCTTACTTCGAATGTCCTGAGGCGAACCCTCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTGTA 270
-----insert pLM1-----
-----ORF pLM1-----
S E L V E K E N K L T C R L E A L N S A H O L O O L P E T I H N I
```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

GCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGSCAGGTCCCTGGATCATCT
CGTCAACCTCCACCTGGACGACTTTCGTCTCTTACTGCGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGACCTAGTAGA

insert pLM1

ORF pLM1

Q L E V D L L K A E N D R L K V A P G P S S G S T P G Q V D G S S

GCATTATCTTCCCCACGCCGCTCCCTAGGCCCTGGCACTCACCCATTCCCTTCGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGAAGGCATCAGTA
CGTAATAGAAGGGGTGCGGGGAGGGATCCGGACCGTGAGTGGGAAGGAAGCCGGGGTCAGAACGCTGTGTCTGGACAGTGGGTACCTACCGTAGTCAT

insert pLM1

ORF pLM1

A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S

CTTGTGGTCCAAAGGAGGAAGTGACCCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCC TGGG
GAACACCAAGGTTTCTCTCTTAC TGGGAGGCCACCACTCTTACGGGGGCGTCTGTAGTAGTTTCCCTGAACTTCGTCGCTCTTAAGAAGGACCC

insert pLM1

ORF pLM1

T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G

CTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTCAAGGACTATATTTCTAAATGGACCCAGCCCTCTACC
GACATCGTTCCAGTCACCTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTTCAAGTTTCTGATATAAAGATTTTACCTGGGTCGGAGATGG

insert pLM1

ORF pLM1

C S K V S G K V D W K M L D E A V F Q V F K D Y I S K M D P A S T

CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAACAGAGTGTGGATGCAGAGCCCCCGAGATGCCCTCTTCCGTCGAGGTG
GACCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACCTTCTCACAACCTACGCTTCGGGGGGCTCTACGGAGGAACGGCAGCTCCAC

insert pLM1

ORF pLM1

L G L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G

TCAATAACATATCAGTCTCCCTCAAAGGTCTTAAGGAGAAATGCGTCGACACCCCTGGTGTTCAGACCGCTGATCCCCAAGCCGATGATGCAGCACTACAT
AGTTATTTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGCAGCTGTGGACCACAAGCTCTGGGACTAGGGGTTCCGGCTACTACGTCGTGATGTAT

insert pLM1

ORF pLM1

V Y N I S V S L K G L K E X C V D S L V F E T L I P K P M R C H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

```
443CC TCCTGCTGAAGCACCGGCGCCTCGTCTCTCGGCCCCAGCGGGCAGGCAAGACC TACCTGACCAATCGCTTGGCCGAGTACCTGCTGGAGCGC 540
TTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTGCGCGTGCCCGTTCTGGA TGGACTGGTTAGCGAACC GGC TCATGGACCACCTCGCG
----- insert pLM1 -----
----- ORF pLM1 -----
S L L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R
TCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTCCAACTAGCCAACCAAG 550
AGACCGGCACCTCAGTGCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCTGCAGAACGTTCC TAGACGTTGACATAGAAAGGTTGGATCGGTTGGTC
----- insert pLM1 -----
----- ORF pLM1 -----
S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N G
TASACCGGGAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC 560
ATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCAC TTCGTCCGAGGTAGTCACTCAACCAAGTTACCCCGGGAGTG
----- insert pLM1 -----
----- ORF pLM1 -----
I D R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T
CTCAAGTATCATAAATGTCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC 570
GACGTTTCATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTT TACTGTGGGTTGGTACCGAACGTGAAC TCGAAGTCC TACAACCTGS
----- insert pLM1 -----
----- ORF pLM1 -----
C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T
TCTCCAAACAGCTGGAGCCAGCCAATGGCTTCCTGCTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAAGAGC 580
AAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCC TCCTTCGACCATCTCAGTCTGTGCTGTAGTTACGGTTGTTCTCTCG
----- insert pLM1 -----
----- ORF pLM1 -----
F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E
TCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCAGCAGCCTCAGACTTCCTCATCGGCCCTTGCTTCT 590
ACGAAGCCACAGAGCTGACCATGGGTTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACCAAGAA
----- insert pLM1 -----
----- ORF pLM1 -----
L L R V L D V V P K L V H L H T F L E K H S T S D F L I G P C F F
```


Page 11

[illegible]

Tuesday, 18 November 1997 11:48

Page 1

fig 34 pLM4 (1 > 10070) Site and Sequence

CATTCAAATATGTATCCGCTCATGAGACAATAACCCGTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAACAGCTGTGGAA
GTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTACGAAGTTATTATAACTTTTCTCTCAGGACTCCGCCCTTCTTGGTCGACACCTT
H S N M Y P L M R Q . P . . M L Q . Y . K R K S P E A E R T S C G 750

TGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
ACACACAGTCAATCCACACCTTTCAGGGGTCCGAGGGGTCTGCTCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGG
M C V S . G V E S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P 770

CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGC
GGTCCGAGGGGTCTGCTCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCS
Q A P Q Q A E V C K A C I S I S Q Q P . S R P . L R P S R P . L R 780

CCAGTTCGCCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGTCAGAGGCGGAGGCCGCTCGGCTCTGAGCTATTCCAGAAGTAGTGAGG
GGTCAAGGCGGGTAAGAGCGGGGTACCGACTGATTAAAAAAATAAATACGTCCTCCGGCTCCGGCGGAGCGGAGACTCGATAAGGTCATCACTCC
P V P P I L R P H A D . F F L F M Q R P R P R P L S Y S R S S E 790

AGGCTTTTTTGGAGGCTAGGCTTTTGAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
TCCGAAAAAACCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTAGCAAGCGTACTAACTTGTCTACCTAACGTGCGTCCAAGAGS
E A F L E A . A F A K I D Q E T G . G S F R M I E Q D G L H A G S P 800

GGCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCCGCTGTTCCGGCTGTCTAGCGCAGGGGCGCCCG
CCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTGTTGCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCCGGGGC
A A V V E R L F G Y D V A Q Q T I G C S D A A V F R L S A Q G R P 810

GTTCTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCGAGCGGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGGCGAC
CAAGAAAAACAGTTCTGGCTGGACAGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACCGGCTC
V L F V K T D L S G A L N E L Q D E A A R L S V L A T T G V P C A 820

CCTGCTCGACGTTGTCACTGAAGCAGGAGGACTGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCTCTCACCTTGCTCTGCGGAGAA
GACACGAGCTGCAACAGTGACTTCGCTCTCCCTGACCGACGATAACCCGCTTACGCGCCCGTCTTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTT
A V L D V V T E A G R D V L L L G E V P G Q D L L S S H L A P A E 830

AGTATCCATCATGGCTGATGCAATGCGCGGCTGATACGCTTGATCCGGCTACCTGCCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGC
TCATAGGTAGTACCGACTACGTTACGCGCGGAGCTATGCGAAGTAGGCGGATGAGCGGGTAAGCTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCTGTC
V S I M A D A M R R L H T L D P A T C P F D H Q A K H R I E R A R 840

ACTCGGATGGAAGCCGCTCTTGCTCATGAGGATGATGAGGAGAGGATCAGGGGCTCGCGCCAGCCGAACGTTTCGCCAGGCTCAAGGCGAGCATGC
TGAGGCTACCTTCGGCCAGAACAGCTAGTCTCTACAGCTGCTTCTGCTAGTCCCCGAGCGGGTGGCTTGACAAGCGGTCCGAGTTCCGCTCTGTA
R M E A G L V C Q D D L D E E H Q G L A P A E L F A R L K A S P 850

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 13

CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCT- 960
GGCTGCCGCTCCTAGAGCAGCAC TGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAGACCTAAGTAGCTGACACCGGCCGA

Kan/Neo
P D G E D L V V T H G D A C L P N I M V E N G R F S G F I D C G R L

GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTCTTTACGG- 970
CCACACCGCCTGGCGATAGTCC TGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCA

Kan/Neo
G V A D R Y Q D I A L A T R D I A E E L G G E V A D R F L V L Y G

ATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC 980
TAGCGGCGAGGGCTAAGCGTCGGTAGCGGAAGATAGCGGAAGAAGTGTCTAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTTGGCTGCGG

Kan/Neo
I A A P D S Q R I A F Y R L L D E F F . A G L W G S K . P T K R R

CAACCTGCCATCAGGAGTTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGSGATGCTCCAGCGC 990
GTTGGAGGTTAGTGTCTAAAGCTAAGTGGCGGCGGAAGATACCTTCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCCGG
P T C H H E I S I P P P P S H K G V A S E S F S G T P A G . S S S A

GGGGATCTCATGCTGGAGTCTTCGCCCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA 900
CCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCTTCTCTGGGCGCGATGCTGCGCTTATT
G I S C V S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .

AAAGACAGAATAAAACGCACGGTGTGGGTGCTTGTTCATAAACCGGGGTTTCGGTCCCAGGGCTGGCACCTCTGTCGATACCCACCGAGACCCCATTS 910
TTTCTGTCTTATTTGCGTGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC
X D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I

GGGCCAATACGCCCCGCTTTCTCTTTTCCCCACCCACCCCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACGTCGGGGCGGCGAGCCCTGCG 920
CCCGGTTATGCGGGCGCAAGAAGGAAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTGGGTTGCAAGCCCGCCGCTCGGGGACGS
G A N T P A F L P F P H P T P Q V R V K A O G S Q P T S G R G A L F

ATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACCTTCATTTTAAATTTAAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACC 930
TATCGGAGTCCAAATGAGTATATGAAATCTAACTAAATTTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAAATATTAGAGTACTGG-
P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . F

AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAAGGATCTTCTTGAGATCCTTTTCTTCGCGTAATCTGCT 940
TTTAGGGAATTGCAC TCAAAAGCAAGTACTCGCACTCTGGGGCATCTTTCTAGTTTCTCTAGAAGAACTCTAGGAAAAAAGAGCGGCATTAGACGAC

pUC ori
K S L N V S F R S T E R Q T P . K R S K D L L E I L F F C A . S A

CTTGCAAAACAAAAAACCCGCTACCAAGCGGTGGTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGGAG 950
GAACGTTTGT TTTTGTGGGCGATGGTGGCCACCAAAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCATTGACCGAAGTCTCTCGGCTC

pUC ori
A C K Q K N H R Y Q P V F V C R I X S Y O L F F R R . L A S A E R F

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 14

ATACCAAATACTGTCTTC TAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA3
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTC TTGAGACATCGTGGCGGATGTATGAGCGAGACGATTAGGACAATGGTC 960

pUC ori
Y Q I L S F . C S R S . A T T S R T L . H R L H T S L C . S C Y Q

TGGCTGCTGCCAGTGGCGATAAGTCGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTS
ACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCAC 970

pUC ori
V L L P V A I S R V L P G V T Q D D S Y R I R R S G R A E R G V R

CACACAGCCAGCTTGGAGCGAAGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGCGGAC
GTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCTCTTTCCGCC TG 980

pUC ori
A H S P A W S E R P T P N . D T Y S V S Y E K A P R F P K G E R R T

AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCTGGTATCTTTATAGTCCGTGTCGGGTTTCGCCACC
TCCATAGGCCATTGCGCGTCCAGCCTTGCTCTCGCGTGC TCCCTCGAAGGTCCCCCTTTCGCGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG 990

pUC ori
G I R . A A G S E Q E S A R G S F Q G E T P G I F I V L S G F A T

TCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTACGGTTCTTGCCCTTTTGCT3
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGTTCGTTCGCCGGAATAATGCCAAGGACCGGAAAAACGAC 1000

pUC ori
S D L S V D F C D A R Q G G G A Y G K T P A T R P F Y G S V P F A

GCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTAT TACCGCATGCAT 10070
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
G L L L T C S F L R Y P L I L V I T V L P P C I

SEQUENCE ID No. 8.

Tuesday, 18 November 1997 10:34

fig 30 pEGFP72 (1 > 9697) Site and Sequence

Enzymes: 72 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Page 1

16p

Bgl I

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCG 100

ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGC

L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

Aat II

CCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT 200

GUGTTGC TGGGGCGGTAAC TGCAGTTATTACTGCATACAAGGGTATCATTGCGGTATCCCTGAAAGGTAAC TGCAGTTACCCACCTCATAAATGCCA

A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

Bgl I

AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTA 300

TTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCATGCGGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCAT

N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

SnaB I

CATGACCTTATGGGACTTTCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGG 400

GTA CTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCCGACCT

H D L H G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V

Aat II

TACGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTG 500

ATGCCAAACTGAGTGGCCCTAAAGGTTGAGAGTGGGGTAAC TGCAGTTACCCCTCAAAACAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT

I A V . L T S I S K S P P H . R Q V E F V L A P K S T G L S K M S .

Nhe I

ACAAC TCCGCCCCATTGACGCAAAATGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCCTAGCGCTG 600

TGTTGAGGCGGGGTAAC TGGGTTTACCGCCATCGGCACATGCCACCCCTCCAGATATATTCGTCTCGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT

Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L

Nco I

CGGGTCGCCACCATGTTGAGCAAGGGCGAGGAGCTGTTCACGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGCGACGTAAACGGGCCACAGTTCAGCG 700

GGGCAGCGGTGGTACCACTGTTCCCGCTCCTCACAAGTGGGCCCCACACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTCAAGTCGC

eGFP.C.e.unc53

F V A T M V S K G E E L F T G V V P I L V E L D G D V H G H K F S

Nco I

TGTCGGCGAGGGCGAGGGGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCAACGGCAAGCTGCCCGTGGCCCTGGCCCAACCTCGTGAC 800

ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCGCTGACTGGGACTTCAAGTAGACGTGGTGGCCGTTCGACGGGCAAGGGGACGGGTTGGGAGCACTG

eGFP.C.e.unc53

V S G E G E S D A T Y S < L T L K F I C T T G K L P V P V P F L V T

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 7

CACCC TGACC TACGGCG TGCAGTGC TTCAGCCGCTACCCCGACCACATGAAGCAGCAGCAGCTTC TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCTGCTGAAGAAGTTCAAGCGGTACGGGCTTCGGATGCAGGTCCCTC 900

eGFP.C.e.unc53

T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

KspI

CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCTTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000

eGFP.C.e.unc53

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAGCCACAACGCTCTATATCATGGCCGACAAGCAGAAGAACGGCATCA
TGAAGTTCCTCC TGCCGTTG TAGGACCCCGTGTTCGACCTCA GTTGTATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTCTTCTGCCGTAGTT 1100

eGFP.C.e.unc53

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I K

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCTGCGCTCGCAGCTCGAGCGGCTGGTGTATGGTCTGCTTGTGGGGGTAGCCGCTGCCGCGGSCACGACGAC 1200

eGFP.C.e.unc53

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGGATCACATGGTCTTGCTGGAGTTCTGTGACCCGCCCGCGGGA
GGGCTGTGGTGTATGGACTCGTGGTCAAGCGGGACTCGTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCCCT 1300

eGFP.C.e.unc53

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

BspM II Bgl II

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTT
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCTTAACCCGGTTAGCGGTGGAAG 1400

eGFP.C.e.unc53

C.e.unc53

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D V A N R H L S

Nru I EcoR I

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAAATGTGATCGTTCCSATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500

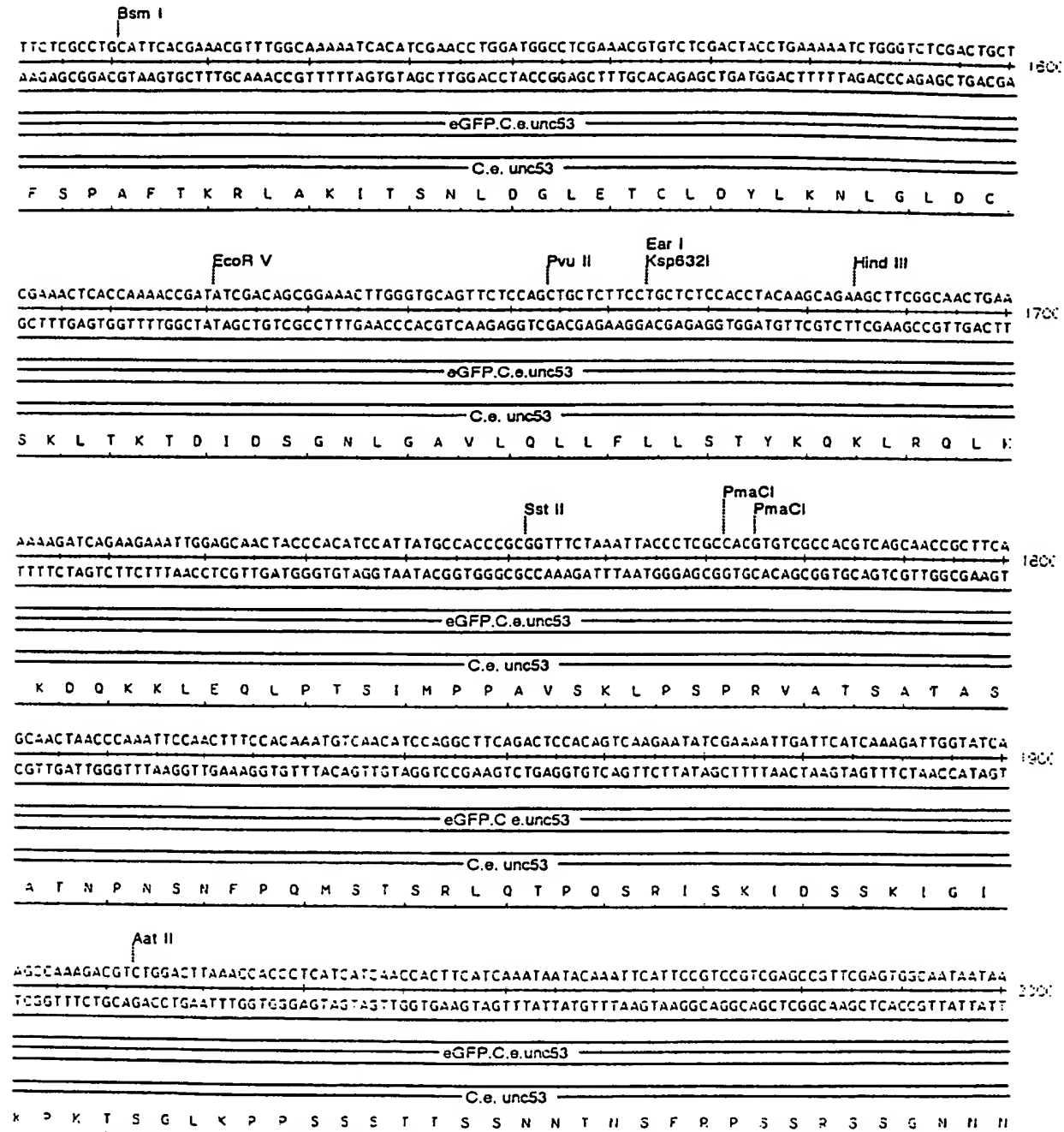
eGFP.C.e.unc53

C.e.unc53

I G S L S K P D I S H D F R D Y R L V S D L I H A I V P I N E

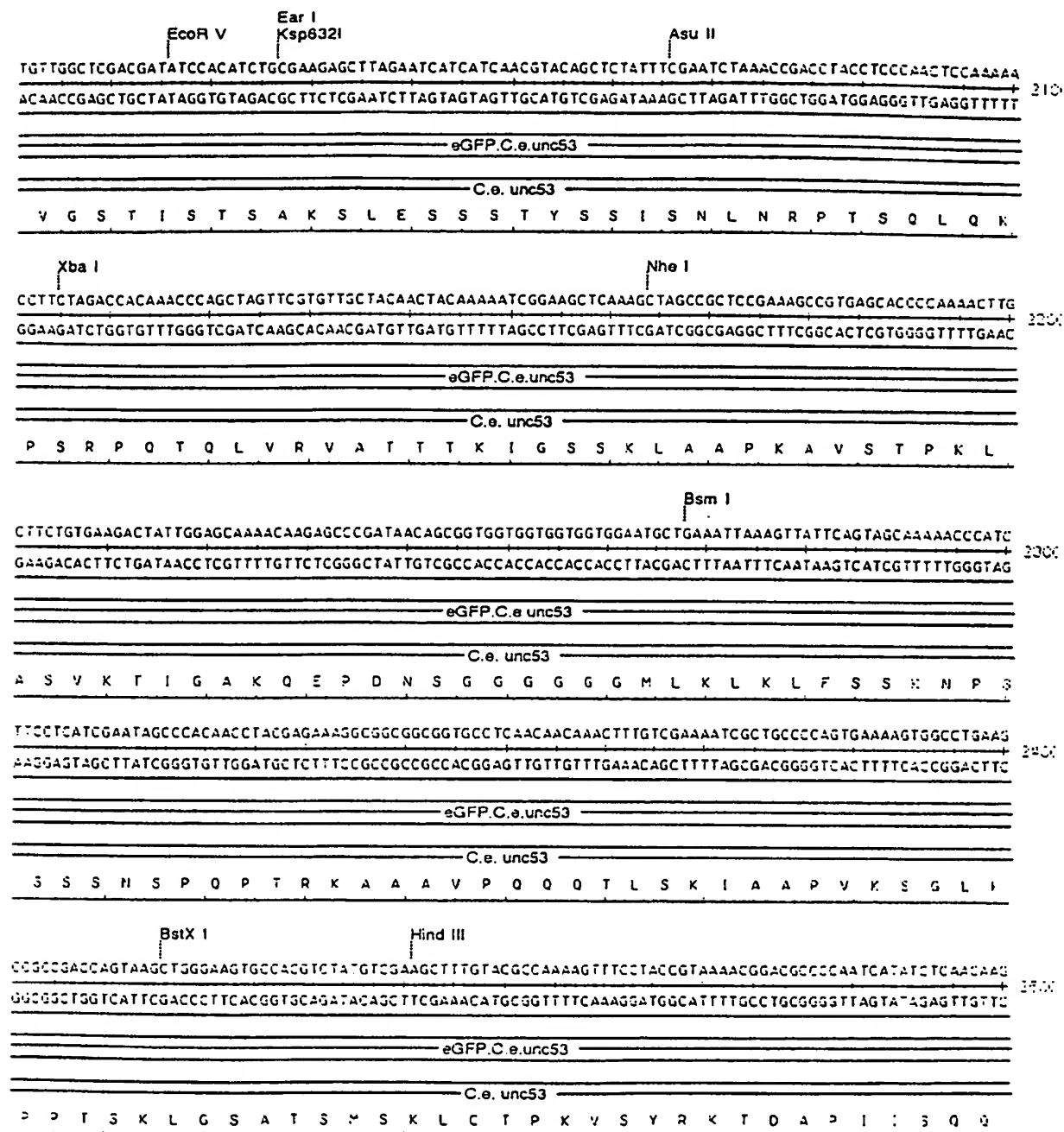
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 3



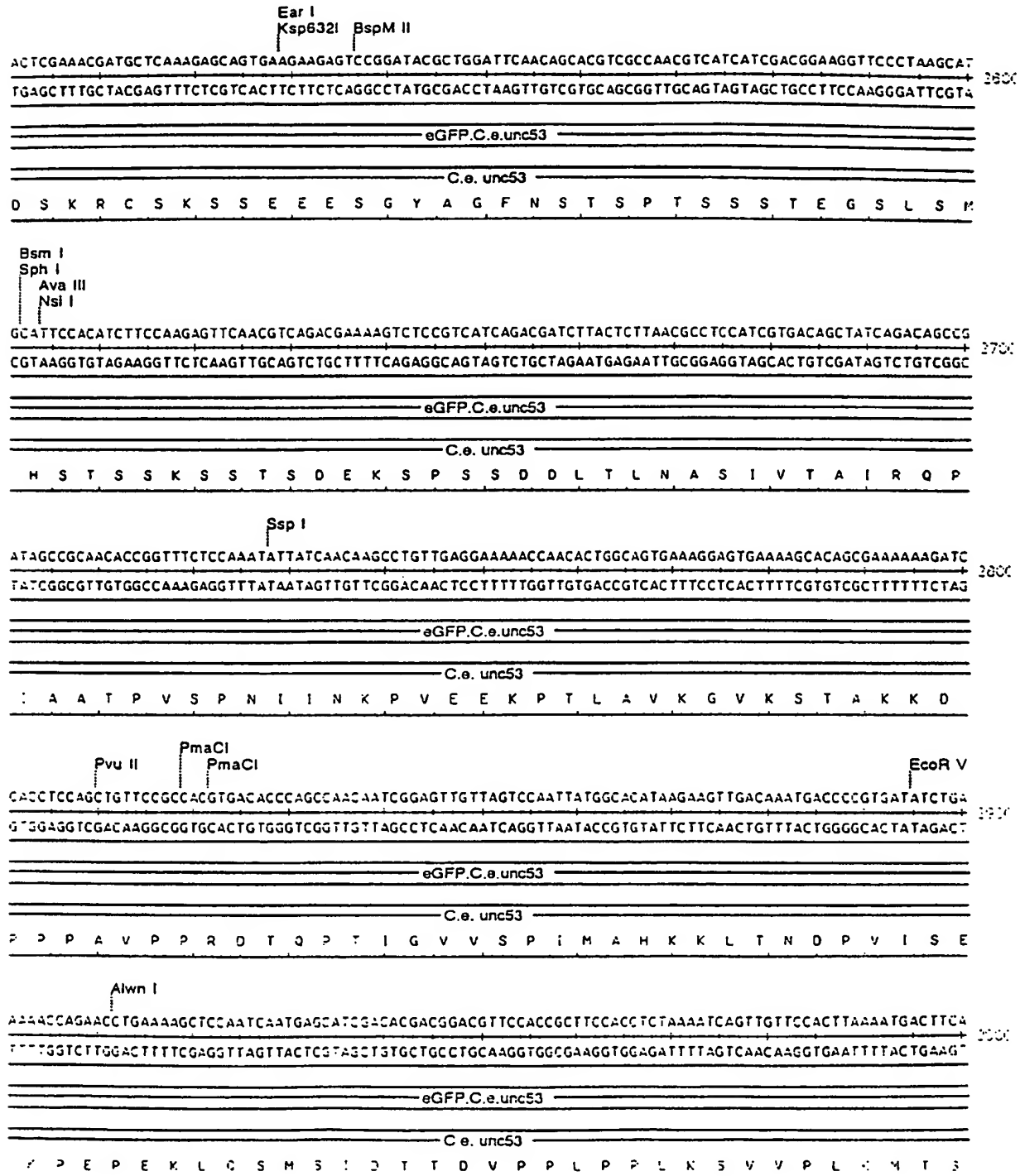
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 6



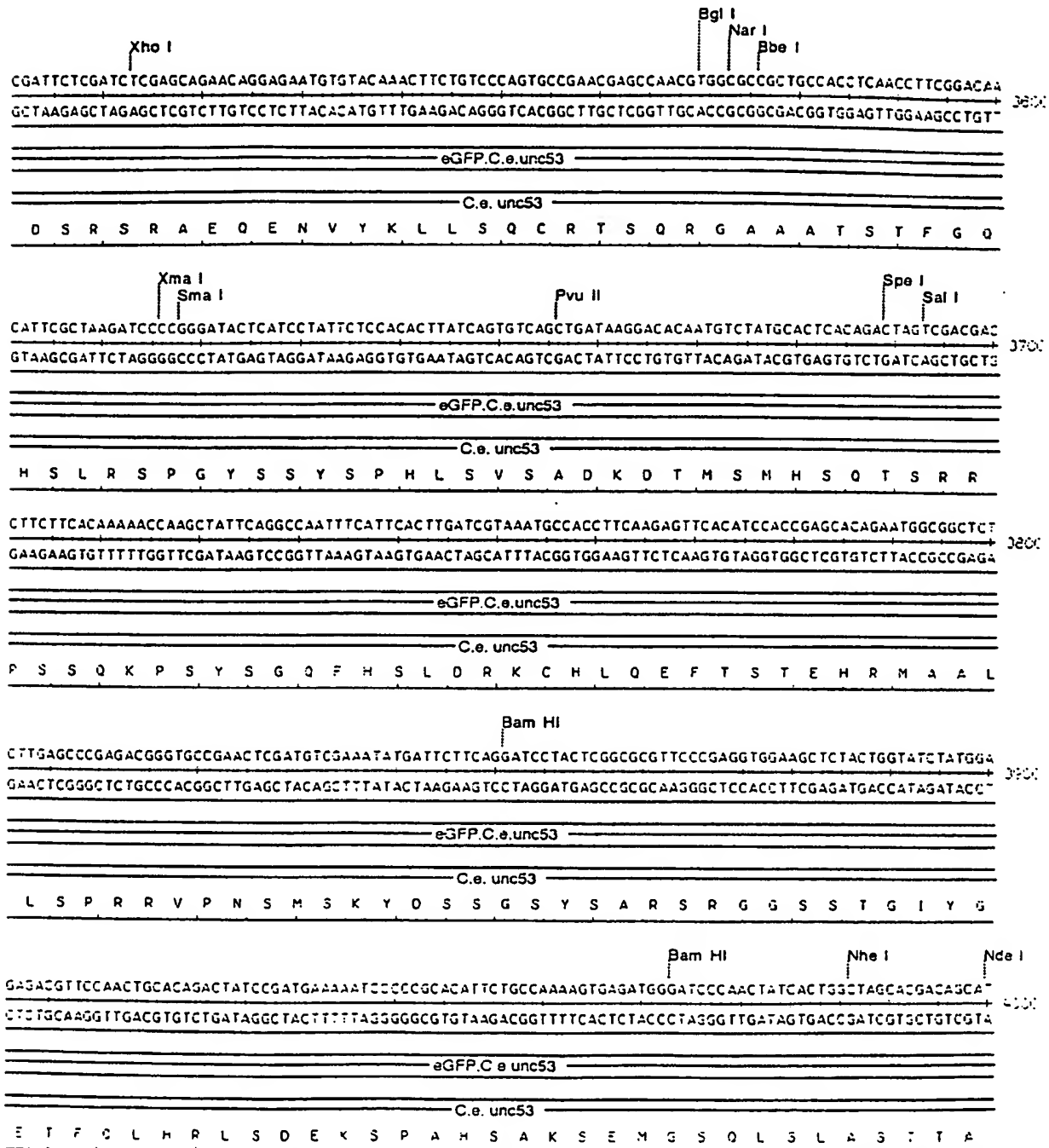
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 6

SpII
SpII
ATCCGACAACCACCAACGTACGATGTTCTTTCTAAAACAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCGCGCTCTGAAGACT
TAGGCTGTTGGTGGTTGCAATGCTACAAGAAGATTTGTTCTTTTATGTTAGCGGACAGTTTCAGCAAACTATACTCGTCAGCAGGCGCAGACTTCTGA 3100
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D
CCATTGTGGCTCATGCTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTCGCTGGAGAGAAGGATGGGAAGAAATAAGACATCAGAATC
GGTAACACCGAGTACGCGCCGAGTCCACTGAGGCGGCTGTTTTGAAGACCATTAGTAAGCGACCTCTCTCTCTACCCTTTCTTATTCTGTAGTCTTAG 3200
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S
CAGCGGCTACACCTCTGACGCCGGTGTTCGATGTGCCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT
GTGCGCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTACTCCCTCTTCGACTTTCCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA 3300
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S G Y T S D A G V A H C A K H R E K L K E Y O D M T R R A O N G Y
Asu II Sst I BspM II
CCTGACAACCTCGAAGACAGTTCTCTCTTGTCTGCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCC TGCTAAACAGGCC TCATCTGTACC 3400
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
P D N F E C S S S L S S G I S D N N E L D O I S T O D L S G V D M
CAACAGTCGCTCCAAACATAGCGACTATTTCCCACTTGTTCGCTATCCACGTCCTTCTTCTCAAAAGCCCGAGTCCCGAGTCGGTCTCCACATCAAT
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAACCAAGCGGTAGGGTGCAGAAGAAGGAGTTTCSGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA 3500
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
A T V A S K H S O Y S H F V R H P T S S S S K P R V P S R S S T S V

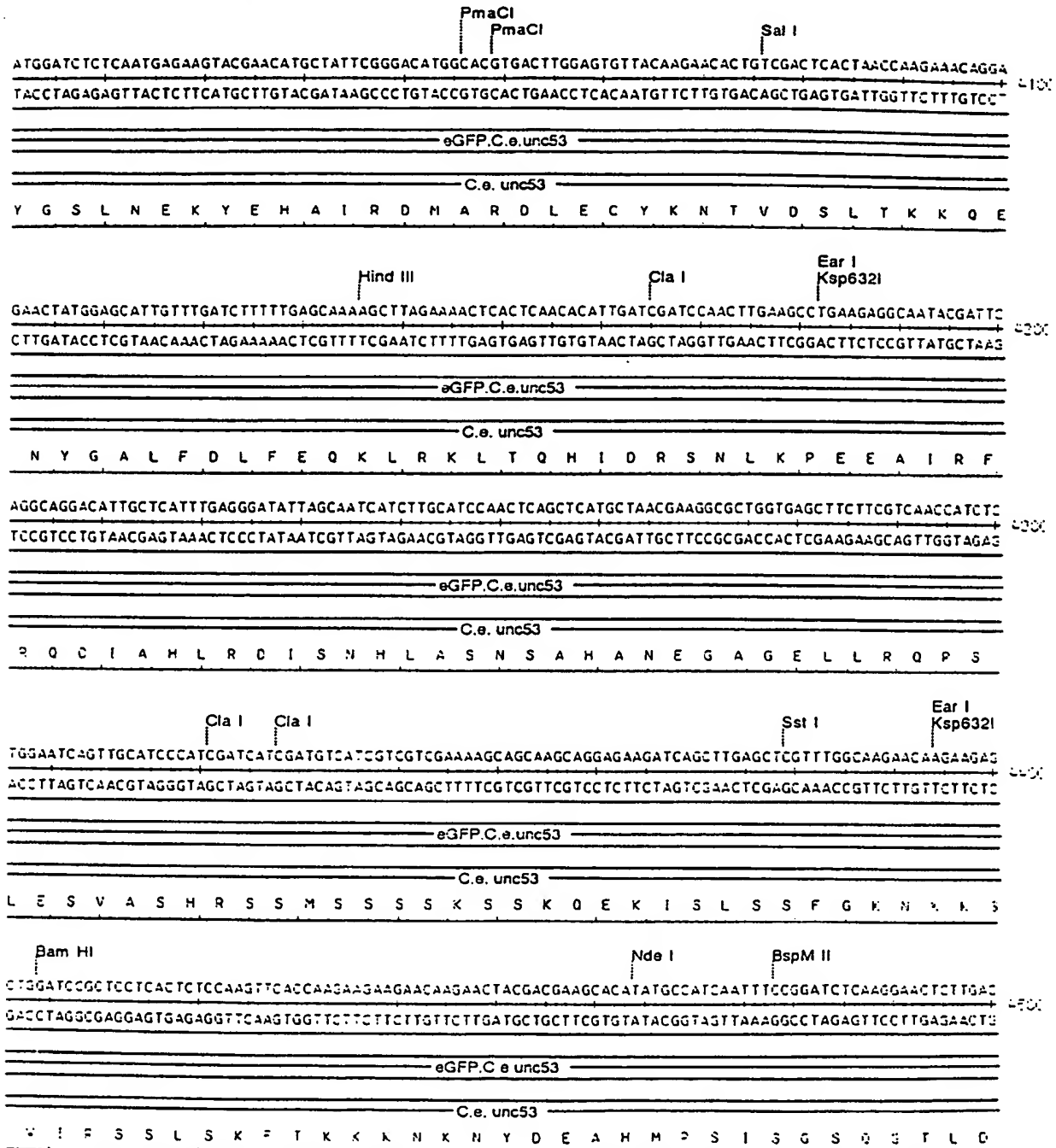
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 7



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 8



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 9

Sst I ApaL I
AACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAAGCGGATAGTGCACITTTACGAAGTCCGCC TTGACAATCTGGATCTGCCCGCCAGGTTGATG 4600
TTGTAAC TACACTAAC TCAACTTCGTTCTCGAGTTTCTTTCGCTATCAGCTGAAATGCTTCAGGCGGAAC TGTAGACCTAGCACGGGCGCTTCAACTAC
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
N I D V I E L K Q E L K E R D S A L Y E V R L D N L D R A R E V D
TTC TGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAAGAAAAGAAGTGGACAAACTCACC AACGGTCCAGGCCACTCGTGC TTTCTTCCCG 4700
AAGACTCCCTCTGTCACTTGTTCACACTTTTGGCTCTTGTTCGTTAATTTCTTTCTTCACCTGTTTGAGTGGTTGCCAGGTCGGTGAGCACG AAGAAGGGC
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
V L R E T V N K L K T E N K Q L K K E V D K L T N G P A T R A S S R
KspI BsrI Asu II
CGCCTCAATTCCAGTTATCTACGACGATGAGCATGCTATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAACGATCCTCTGGCTGCAAC 4800
GCGGAGTTAAGGTCAATAGATGCTGCTACTCGTACAGATACTACGTCGCACATCGTCATGATGATCAGTTAGAAGCTTTGCTAGGAGACCGAGCTT
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
A S I P V I Y D D E H V Y D A A C S S T S A S Q S S K R S S S C H
Pvu I Hpa I EcoR V
TCAATCAAGGTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTCGATCGTTAACC CGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCA 4900
AGTTAGTTCCAATGACATTTCACCTGTAGCGACCTCTTTAGTCAAGCTAGCAATTGGGCTGTTTCTCTATTAGCATCCTATAGAACGGTACAGTTGG
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
S I K V T V N V D I A G E I S S I V N P O K E I I V G Y L A R S T
Cla I
GTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGSACTATTTGAAGTCTACCTATCCAGAAATGATGTGGAGCATCAACTTGGAAATCGATGCTCC 5000
CAGTCAGTACGACCTTTCTGTAAGTACAAAGATAAGATCTGTATAAATTCAGATGGATAGGCTTTAACTACACCTCGTAGTTGAACCTTACCTACGAGC
-----eGFP.C.e.unc53-----
-----C.e.unc53-----
S Q S C W K D I D V S : L G L F E V Y L S R ! D V E H Q L G I D A F

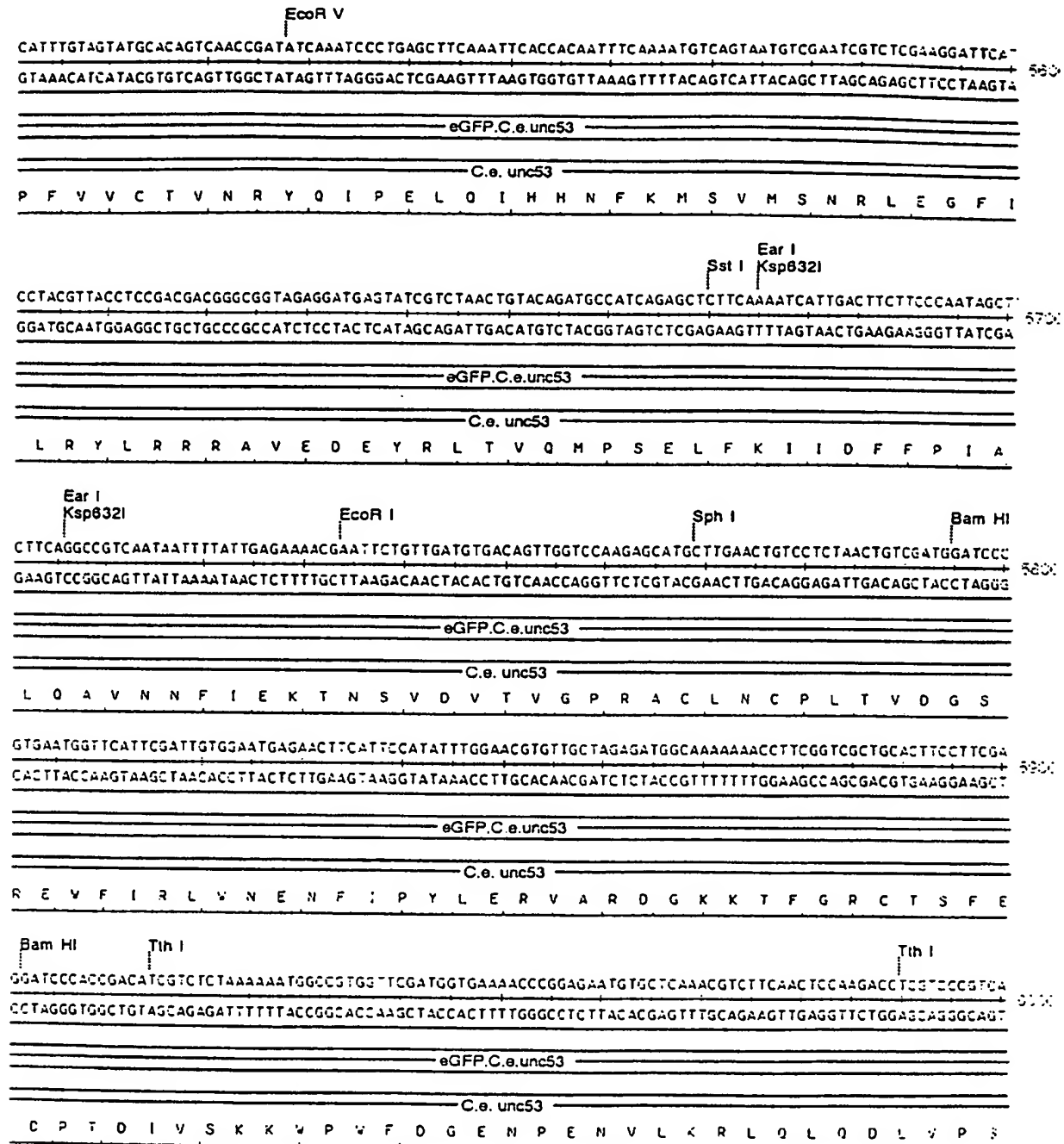
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 10

Mlu I
TGATTCTATCCTTGGCTATCAAATTGGTGAAC TTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAGCCATCCAAC TGACATTCTTACTTCCTCA 5100
ACTAAGATAGGAACCGATAGTTTAACCACTTGAAGCTGCGCAGTAACCTCTGAGGTGTTGGTACTATTGGTCGGTAGGTTGACTGTAAGAATGAAGGAG
eGFP.C.e.unc53
C.e. unc53
D S I L G Y O I G E L R R V I G D S T T M I T S H P T D I L T S 3
ACTACAATCCGAATGTTTCATGCACGGTGCCGACAGAGTCGGTAGACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAATGATTCTCCAACTCGTCA 5200
TGATGTTAGGCTTACAAGTACGTGCCACGGCGTGCTCAGCGCATCTGTGACAGCAGGAACATACGAAGAAGGTTTCGTTTACTAAGAGGTTGAGCAG
eGFP.C.e.unc53
C.e. unc53
T T I R M F M H G A A O S R V D S L V L D M L L P K Q M I L Q L V
Aat II BsrI BsrI Asu II
AGTCAATTTTGACAGAGAGACGCTCTGGTGTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCCTGGCTGCTTATGTATCTATTGGAAC 5300
TCAGTTAAACTGTCTCTGACAGACCACAATCGACCTCGTTGACCTTAACTTTCTGTTGACCGCTTCTGGGACCGACGAATACATAGATAAGCTT
eGFP.C.e.unc53
C.e. unc53
K S I L T E R R L V L A G A T G I G K S K L A K T L A A Y V S I R T
Bsm I Xmn I Bgl II
AAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAGAAGAATTGCTTCAAGTGGAAACGACGCCCTGGAAAAGATCTTGAGAGGC 5400
TTTAGTTAGGCTTCTATCATAACAATTATAGTCGTAAGGACTTTGTTATTTCTTCAACGAAGTTCACCTTGCTGCGGACCTTTTCTAGAACTCTTCG
eGFP.C.e.unc53
C.e. unc53
N Q S E D S I V V I S I P E N N K E E L L Q V E R R L E K I L R S
Ava III Nsi I Xba I
AAAGAATCATGCATCGTAATCTAGATAATATCCCAAAGAAATCGAATTGCATTGTTGTATCCGTTTTTGCAAAATGTCCCACTTCAAAACAACGAAGGTC 5500
TTCTTAGTACGTAGCATTAAAGATCTATTATAGGTTTCTTAGCTTAACGTAAACAACATAGGCAAAAACGTTTACAGGGTGAAGTTTTGTGCTTCCAG
eGFP.C.e.unc53
C.e. unc53
K E S C I V I L C N I P K N R I A F V V S V F A N V P L Q N H E G

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 11



Page 12

BspM I | **Xho I** | **Sph I**

CCTGCCAACTCATCCCGACAACTTCAATCCCCTCGAGTCGTTGATCCAATTGCATGTACCAAGCATCAGACCATCGACAACATTGAACAGAAGACT
GGACGGTTGAGTAGGGCTGTTGTGAAGTTAGGGGAGCTCAGCAACTAGGTTAAGTACGATGGTTCGTAGTCTGGTAGCTGTTGTAACCTTGCTTCTGTA

eGFP.C.e.unc53

C.e.unc53

P A N S S R Q H F N P L E S L I Q L H A T K H Q T I D N I . T E D

Asp 718
Kpn I

CTAATCTTCTCTCGCCTCTCCCCCGCTTTTCCTTATCTTCGTACCGGTACCTGATGATTCCCCATTTTCCCCCTTTTCCCCCAATTTCCCAGAACCCTCT
GATTAGAAGAGAGCGGAGAGGGGGCGAAAGGAATAGAAGCATGGCCATGGACTACTAAGGGGTAAAAGGGGGAAAGGGGGTTAAAGGGCTTTGGAGGA
S N L L S P L P R F P Y L R T G T . . F P I F P L F P P I S Q N L L

Xma I | **Sma I** | **Dra I** | **Xmn I**

GTTCCCTTTGTTCTAGTCTCTCCGGGTGCCGACGCCGAAGCGATTAAAAACCTTTTTCTTTCGAAACATTTCCTATTGCTCATAATAGTCAAATTG
CAAGGGAACAAGGATCAGGAGGGCCACGGCTGCGGCTTCGCTAAATTTTGGAAAAAGAAAGGCTTTGTAAAGGGTAACGAGTAATTATCAGTTTAAC
F P L F L V L P G A D A E A I . K P F S F R N I S H C S L I V K L

Xma I
Apa I
Sma I
Xho I | **Bam HI** | **Xba I** | **Bcl I**

AATAAACAGTGTATGTACTTAAAAAAAAAAAAAAAAAACTCGAGGGGGGGCGCGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCCACA
TTATTGTACATACATGAATTTTTTTTTTTTTTTTTTTGAGCTCCCCCGGGCCCTAGGTGGCCTAGATCTATTGACTAGTATTAGTCGGTATGGTGT
N K Q C M Y L K K K K K K L E G G P G I H R I . I T D H N Q P Y H

Dra I | **Bsm I** | **Hpa I**

TTGTAGAGGTTTTACTTTGCTTTAAAAAECTCCACACCTCCCCGTAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTCAS
AAACATCTCCAAAATGAACGAAATTTTTGGAGGGGTGGAGGGGGACTTGGACTTTGTATTTTACTTACGTTAAACAACAATTGAACAAATAACGTC
I C R G F T C F K K P P T P P P E P E T . N E C N C C C . L V Y C S

Bsm I

CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTCACTGCAATCTAGTTGTGGTTTGTCCAAACTCATCAATGT
GAATATTACCAATGTTTATTTTCGTTATCGTAGTGTTTAAAGTGTTTATTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACA
L . V L O I K O . H H K F H K . S I F F T A F . L W F V Q T H Q C

Mlu I | **Ssp I**

ATCTTAACGCGTAATTTGTAAGCGTTAATATTTTGTFAAAATTCGCGTTAAATTTTTGTFAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGGAA
TAGAATTGGSCATTTAACAATTCGAATTTATAAACAATTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAGAAATTTGGTTATCCGGCTTTAGCGGTT
! L T R K L . A L I F C . N S R . I F V X S A H F L T N R P K S A

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 13

Bsr I

AATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTCAAA
TTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAGGTCAAACCTTGTTCACGGTGATAATTTCTTGCACCTGAGGTTGCAGTTT
K S L I N Q K N R P R . G . V L F Q F G T R V H Y . R T V T P T S I 6800

Dra III

GGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACC
CCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGTATTTAGCCTTGG
G E K P S I R A M A H Y V N H H P N Q V F W G R G A V K H . I G T 6900

Nae I

CTAAAGGGAGCCCCGATTAGAGCTTGACGGGGAAAGCCGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGC
GATTTCCCTCGGGGGCTAAATCTCGAACTGCCCTTTTCGGCCGCTTGACCCGCTCTTTCCCTTCCCTTCTTTCGCTTTCTCGCCCGCATCCCGCGACCG
L K G A P D L E L D G E S R R T V R E R K G R K R K E R A L G R V 7000

Ksp I

AAGTG TAGCGGT CACGCTGCGGTAAACCACACACCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGA
TTCACATCGCCAGTGGCAGCGCATTTGGTGGTGTGGGCGCGCGAATTACGCGCGCATGCCCGCGCAGTCCACCGTGAAAAGCCCCCTTACACGCGCCT
Q V . R S R C A . P P H P P R L M R R Y R A R Q V A L F G E M C A E 7100

BspH I

ACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTC
TGGGATAAAACAAATAAAAAGATTATGTAAGTTTATACATAGCGGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTCTTCTCAS
P L F V Y F S K Y I Q I C I R S . D N N P D K C F N N I E K G R V 7200

Sph I
Ava III
Nsi I

CTGAGGCGGAAGAACCAGCTGTGGGAATGTGTGTCATTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGAGGAGAGTATGCAAGCATGCATCTCAAT
GACTCCGCTTTCTTGGTGCACACCTTACACACAGTCAATCCACACCTTTTCAGGGTCCGAGGGGTCGTCGCTCTTCATACGTTTCGTACGTAGAGTTA
L R R K E P A V E C V S V R V V K V P R L P S R Q K Y A K H A S Q 7300

Sph I
Ava III
Nsi I

TAGTCAGCAACCAGGTGTG3AAAGTCCCCAGGCTCCCCAGCAGGAGAGATGCAAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAA
ATCAGTCGTTGGTCCACACCTTTTCAGGGGTCCGAGGGTTCGTCGCTCTTCATACGTTTCTACGTAGAGTTAATCAGTCGTTGGTATCAGGTCGGGGATT
L V S N Q V V K V P R L P S R Q K Y A K H A S Q L V S N H S P A P H 7400

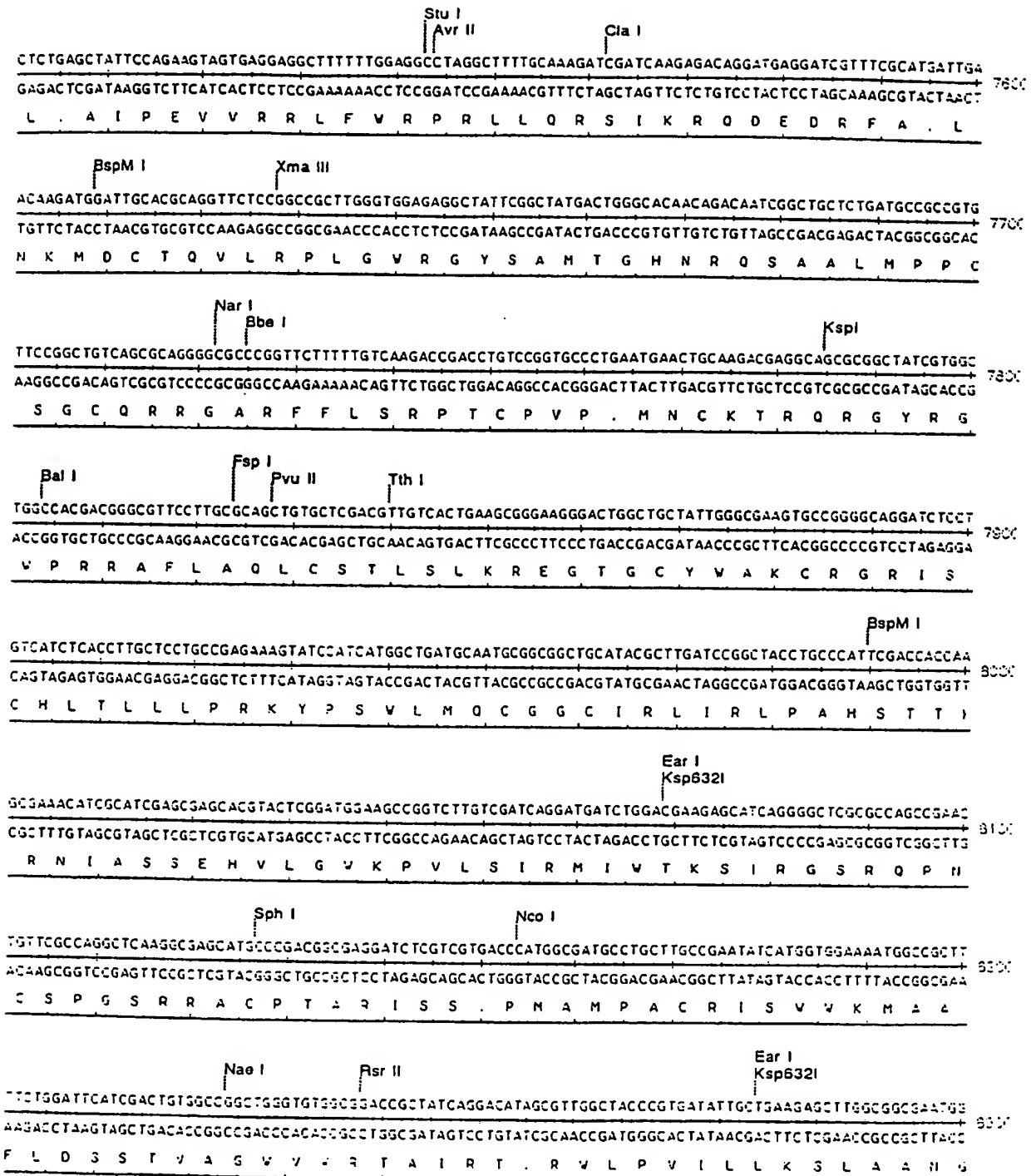
Bsr I

CTCCGCCCATCCCGCCCTAACTCCGCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAAGAGCCGAGGCGGCTCGG
GAGGCGGGTAGGGCGGG3ATTGAGCGGGTCAAGCGGGTAAGAGGCGGGTACCGACTGATTAATAAATAAATACGTCCTCGGCTCCGCGGAGGCGG
S A H P A P N S A Q F R P F S A P V L T N F F Y L C R G R S A L G 7500

Bgl I
Sfi I

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9897) Site and Sequence

Page 14

GCTGACCGCTTCCTCGTGTTCACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCTTCTATCGCCCTTCTTGACGAGTTCCTTGAGCGGGACTCTGGG
CGACTGGCGAAGGAGCAGCAATGCCATAGCGCGGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCC 840
L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G

Asu II BspM I
GTTGGAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGAGATTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTCCGG
CAAGCTTTACTGGCTGGTTCGCTCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGCGGAAGATACTTCCAACCCGAAGCCTTAGCAAAAGGCC 850
V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F P

Nae I KspI Avr II
GACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCGCCACCCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCG
CTGCGCGCGACCTACTAGGAGGTGCGGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCC 860
G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G

KspI
AAGGAACCCGCGCTATGACGGCAATAAAAGACAGAATAAAACGCACGGTGTGGGTGCTTTGTTTCATAAACCGGGGTTGGTCCCAGGGCTGGCACTC
TTCCTTGGGCGGATACTGCCGTATTTTCTGTCTTATTTTGGGTGCCACAAACCAGCAACAAGTATTTGCGCCCAAGCCAGGGTCCGACCGTGAG 870
R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L

GTGATACCCACCGAGACCCATTGGGGCAATACGCCCGGTTTCTTCTTTTCCCCACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCA
ACAGCTATGGGTGGCTCTGGGTAACCCCGGTTATSCGGGCGCAAAGAAGGAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCGAGCGT 880
C R Y P T E T P L G P I R P R F F L F P T P P K F G . R P R A R

AluI OxaN I Dra I Dra I
GCCAACGTCGGGGCGGAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGA
CGGTTGCAGCCCGCGCTCCGGACGGTATCGGAGTCCAATGAGTATATAGAACTAACTAAATTTTGAAGTAAATTTAAATTTCTAGATCCACT 890
S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E

BspH I
AGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA
TCTAGGAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACCTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTAGAAGAACTCT 900
D P F . . S H D O N P L T . V F V P L S V R P R R K D Q R I F L R

TCTTTTTTCTGCGGTAATCTGCTGCTTGCAACAAAAAACACCGCTACCAGCGGTGGTTTGTITGCGGATCAAGAGCTACCAACTCTTTTCCG
AGGAAAAAGACGCGCATTAGACGACGAACGTTTGTITTTTGGTGGCGATGGTCCGACCAACAAACGGCCTAGTTCGATGGTTGAGAAAAAGGC 910
S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S

BsrI
AAGGTAAGTGGCTTCAGCAGAGCGCAGATACAAA*ACTGTCTTCTAGTGAGCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT
TTCATTGACCGAAGTCTGCTCGCTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGTTGGTGAAGTTCTTGAGACATCGTGGCGGATGTA 920
E G N V L Q O S A D T < ' C P S S V A V V R P P L Q E L C S T A Y I

Tuesday, 18 November 1997 10:35
fig 30 pEGFP2 (1 > 9697) Site and Sequence

Page 16

Nhe I

ACCTCGCTCTGCTAATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGT
P R S A N P V T S G C C Q V R . V V S Y R V G L K T I V T G . G A

Apa I

GCGGTCGGGCTGAACGGGGGTTCTGTCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC
CGCCAGCCCGACTTGCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACCTCGATACCTCTTCGCGG
A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K R

ACGCTTCCCGAAGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATC
TGCGAAGGGCTTCCCTCTTTCGCTGTCCATAGGCCATTTCGCGTCCCGAGCCTTGCTCTCTCGCGTGTCTCCCTCGAAGGTCCCCCTTTGCGGACCATAG
H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V S

TTTATAGTCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC
AAATATCAGGACAGCCCAAGCGGTGGAGACTGAATCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGCGTGTGCGCCG
L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R G

Ava III
Nsi I

CTTTTACGGTTCTTGGCCTTTTCTGTCGCTTTTCTGCTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
GAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Monday, 1 December 1997 14:12
fig pCB501

SEQUENCE I.D. No 9

Page /

	10	20	30	40	50	60	70	
A T G A C C A T G A T T A C G C C A A G C T T G C A T G C C T G C A G G A A T T C G A T A T C A A G C T T A T C G A T A C C G T C G A C C T							/ 0	
G A G G A T C A G A A G A A A T T G G A C C A A C T A C C C A C A T C C A T T A T G C C A C C G C G G T T T C T A A G T G A G T T G A A							1 / 0	
T T T T G A G T T A C G A C T A C A A A A T G T G T T T T A A A C T A T C T T C G A C T T G A G T C T A T T C T G T A T G A C T							2 / 0	
A G T T G T T G A G T G A T T T T C A T T G A G A A A T A T T A A A A G G A A C A T T A T T A C T T T G C T G A T T T G C C C T A A C							280	
T T T G A T T T A G T T T T T C S A T C A A C T A G A T C T T A C A A A A C T T G C A A T A C A A T T C C A T T T T C A G A T T A C C C T C							360	
	360	370	380	390	400	410	420	
S E C A C G T G T C G C C A C G T C A G C A A C C G T T C A G C A A C T A C C C A A A T C C A A C T T C C A C A A A T G T C A A C A							420	
T C C A G G C T T C A G A C T C C A C A G T C A A G A A T A T C G A A A A T T G G T A A G A A T T T T A G T T T G A G C T C A A A C T T G T							480	
A T A A A A T G C C C A G A A A A G A A G A T G A T A A A A A T G T A G T T T T T T G C A A A A C T T C C A C C T T T A T T G C T C T A A							560	
T A T G A C G G C T T A T A T C T C A A T T T T C T T G A G T T T A T C A A A A A A T T T T G C C A C T A T A C A A A G T A G A A A A G T							630	
A T T T G C A C A A A T T T T S T C A G T T G A C A C C T T T G T A A T A G A T C C A A A T G G A A C C T A G A T A C A A G C T G T T A A							700	
	710	720	730	740	750	760	770	
A S T G G A A G G A G C G C A A G T C T A T A C T S E A A A F A A T G A T C T G A A A C A A A T T T G T G C T A T T C T C A A A T G T T T A							/ / 0	
A B A C A T S T T T T T S A A G A T T T T T T C A A A T T C G C A C T A G T T T C A G A A C C T T C C T T T T T G T A T C A A A A A G T A A A							840	
A A A A A A C T A T T T T C A A A C C C T C A C C G C C A C C A T G T T T C A A C T C T T A A T T T T T A T A A A T T T T G C A A T T T T A C							910	
A A A T C G C C T C C C C T T G C C C G A A A A G T C C C A C C A A A A A T C A A T T T C G G C T C A T A A T G A C T T T T A A A T T							980	
C A T G T G A G A A A A C A C A G A A G A G G C T A A T A A A T T G A C A G G A C A G G T T G T G C C T C T T C C C C C C C T C T C							1050	
	1060	1070	1080	1090	1100	1110	1120	
C G C C T C C T C C T C C C G T T C C A T C T C C A A C A C A A C A A T T T T C C A A T T T C G T T G T C C A T T T T G C T T A G A A A							1120	
C A T T T G T G T G T G G A A A G A A A C T A C A C C G G A G A C G G T C A A T T A A T T C G A A T G A G A G A T G G G A A T T A C T C							1190	
T T T C G G A A A T T G A T G A A A A A G A T A C A C C G A T G A C A C T G G C T G G T A G T A G T A G T G A A T T G A A T T G C T T							1260	
T T C A T C G T C T C A A C T T G C G C A T G A G T T T T C C C G C T C T C A T C A C T G A C A A T T A A T G T C G G G T T T T A T G							1330	
C G C T C T T T C C T A T T C C G C A C T C A T T C T G S S T T A C C A C A A A C T G G A A T A C A T T T A C T A C T A T T C A A G C C							1400	
	1410	1420	1430	1440	1450	1460	1470	
A T T A T T T T T C A T A T T F A A T T T T S T G C A A T T A G G G A T A A A C A C G A C T T T T A A A A G T T A T T T A A A A A A C G							1470	
A T A T T T T C G A T T T T A A A A A T C T G A A A A G T T C A A A A A A T C A A T A A A T A T T C C C T A A C A A A T T G T A T G C C							1540	
I A A A A I I I A A T T C T A C T G T T G A C A A T A T C T T T A T A I S I A T C A C T G T T T T C C A T C T C A A A A C C T T G A A T C							1610	
C C C C A A G T T A T A G G A A C C T C C G T G T C A C A T T T C C C A T G C T A T G A A T G C T A C T C A G C A C A T A T C C A A A A							1680	
T T A A G C T A G A C G G T T S A T A A T T A T T G S S C A G S C G T A A T A A A G T G C A A G C A G T T A G A A T T T T A A T T C A A G C							1750	
	1760	1770	1780	1790	1800	1810	1820	
A B A A T T A T C T A T C A A A T T C A A T C T T T C A A C A T T C A C C C A G T T C G T A C A A T T T C C A T G C T T T T T G G C C							1820	
A T T A A A A A A C T T T C T C A C C C C T T C A T C C T C T C A C T C G T A T C A T A A A A A G T A T A G C A A A A G C C C G A C T C T							1890	
A C T T T T T A A G A G A A A G A G A T A C T G A G C C A C A T G G C G T G T G A C C C T T T C A T C T C G T C C G T T C G G C T C A A							1960	
A T T C A C G C T C A T A C T A A C T C T T C A A T A B C C A T A G A C C T C C T T G T T T C T C T T C G T T T T G A C T C C C C C							2030	
T A T T T T T S T G G C T G C C T G A A A G C C G S B A A A T T A G T A T A T T A T G A C C T A T C T T T A T G C A A T A C A T A							2100	
	2110	2120	2130	2140	2150	2160	2170	
A A A A A C G A G S C A A T T A A A A A T A T T A A A T T A A T G A G G T T G T A G A T G T A G A T T T G G A A A A G A A A A A A A							2170	
A T G A A C A A A A T A G G A A C C G C C A G A T C A A A A T T C T A T T T A A A A G T T T C A A A								

Monday, 1 December 1997 14:12
lg pCB501

Page 2

```
2460      2470      2480      2490      2500      2510      2520
AAATGTTTCCAGTTTCCTGTTACCTAAAATCTTAATGCCCTGTGATGCGTAAAATCGTTATCCCTTTCTC 2520
ACACCTTTCATTTACAGATTTCATCAAGATTGGTATCAAGCCAAAGACGCTGGACCTAAACCACCCCTC 2590
ATCATCAACCACCTTCATCAAAATACAAAATCATTCGGTCCGTCGAGCCCTTCGAGTGGCAATTAATTA 2660
GTTGGCTCGACCATATCCACATCTGCGAAGAGCTTAGGTATCCGATCCTTCGGCTCTTTTATAGAAAT 2730
ATATTATTTTCAGAAATCATCATCAACGTACAGCCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800

2810      2820      2830      2840      2850      2860      2870
AAAAGCTCTTAGACCACAAACCGAGCTAGTTCTGTGTTGCTACAACACAAAAATCGGAAGCTCAAAGCTA 2870
CCGGCTCCGAAGCCGTGAGCAGCCCAAACTTGTCTCTGTAAGACGATGGAGCAAAACAGAGCCCG 2940
ATAACAGCGGTGGTGGTGGTGGTGGTAAATGCTGAAATTAAGTTATTCAGTAGCAAAAACCCAGCTTCT 3010
ATCGAATAGCCACAACTACGAGAAAAGGCGGGCGGTGCTCAACAACAACTTTGTCGAAAAATCGCT 3080
TCCCGAGTGAATAATGGCTGAAGCCCGACCAAGTAAGCTGGGAAGTGCCACGCTATGTGGAAGCTTT 3150

3160      3170      3180      3190      3200      3210      3220
ATACGCCAAAAGTTTCTTACCGTAAACCGGACGCCCAATCATATCTCAACAAGACTCGAAACGATGCTC 3220
AAAGAGCAGTGAAGAAGAGTCCGGATACGCTGCATTCAACAGCAGCTCGCCAACGTCATCATCGACGGAA 3290
CTTCCCTAAGCATGCATTCCATATCTTCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCTATCAGACG 3360
ATCTTACTCTTAACGCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTCTCCAAA 3430
TATTATCAACAAGCTTTTGAAGAAAACCAACACGCGCAGTGAAGGAGTGAATAACAGCGGAAAAAA 3500

3510      3520      3530      3540      3550      3560      3570
ATCCACCTCCAGCTGTTCCGGCCACGTTGACACCCAGCCAACTCGGAGTTGTTAGTCCAATTATGGCAC 3570
ATAAGAAGTTTACAAATGACCCCGTGATATCTGAAAAACAGAACCTGAAAAGCTCCCAATCAATGAGCAT 3640
TGACAGGACGGAGCTTCCACCGCTTCCAGCTCAAAAATCAGTTGTTCCACTTAAAAAGACTTCAATCCGA 3710
CAACCCACCAACCTACGATGTTCTTCAAAAAGGAAAAATCAGATCGGCTGTCAASTCGTTTGGATATG 3780
AGCAGTCTGTCGGCTCTGAAGACTCCATTGTTGGCTCATGCTCGGCTCAGGTGACTCCGCCGACAAAAAC 3850

3860      3870      3880      3890      3900      3910      3920
TTCTGTTAATCATTTCCGTCGAGAGAAAGGATGGGAAAAGAATAAGACATCAGAAATCCAGCGGCTACACCTCT 3920
TAGCCCGGCTGTTGCGATGTTGCGCCAAAATGAGGGAGAAAGCTGAAAGAATACGATGACATGACCTCGGAG 3990
CACAGAACGGCTATCTGACAACTTCAAGAGAGTTCCTGCTTGTCTCTGGAATATCCGATAACAACGA 4060
GCTCGACGACATATCCACGGAGGATTTGTCGGGAGTAGACATGGCAACAGTCCGCTCCAAACATAGCGAC 4130
TATTCCCACTTTGTTCCCAACGAGCTCTTCTGCTCAAAAGCCCGAGTCCCCAGTGGTCTCCACAT 4200

4210      4220      4230      4240      4250      4260      4270
CAGTCCATTCTCGATCTCGAGCAGAAACAGGAGAAATGTGTACAAACTTCTGTCCAGTCCCGAACGAGCCA 4270
AGCTGGCGCCGCTGCCACCTCAACCTTCGGACAACTTCGCTAAGATCCCGGGGATACCTCATCTATCTCT 4340
CCACACTTATCAGTCTCAGCTGATAAGGACCAATGTCTATCCACTCACAGACTAGTGGACGACCTCTCT 4410
CCCAAAAACCAAGCTATTCAAGGCCAATTTCACTTGTATCGTAAATGCCACCTTCAAGAGTTCACATC 4480
CAGCGACACAGAAATGGGGGCTCTCTTGAACCCGAGACGGGTGCCGAACTCGATGTCGAAATATGATCTCT 4550

4560      4570      4580      4590      4600      4610      4620
TCAGGATCTTCTCGATCTCGAGCAGAAACAGGAGAAATGTGTACAAACTTCTGTCCAGTCCCGAACGAGCCA 4620
AGCTGGCGCCGCTGCCACCTCAACCTTCGGACAACTTCGCTAAGATCCCGGGGATACCTCATCTATCTCT 4690
TAGCAGGACAGCATATGGAATCTCTAATGAGGAAGTACGAACATGCTATTCGGGACAAGGACCTCACTTG 4760
GAGTGTACAGAAACATGTCGACTCATTAAGCAAGAAACAGGAGAACTATGAGGATCTCTTGAATCTCT 4830
TGAAGCAAAAGCTTAGAAAACCTCATTCATCAATGATCGATCCAACCTGAAAGCCCTCAAGAGGCAATACG 4900
```

Monday, 1 December 1997 14:12
fig pCB501

Page 3

4910 4920 4930 4940 4950 4960 4970
A T C A G G C A G G A C A T T G C T C A T T T G A G G G A T A T T A G C A A T C A T C T T G C A T C C A A C T C A G C T C A T G C T A A C 4970
G A A G G C G C T G G T G A S C T T C T T C G T C A A S C A T C T C T G A A T C A G T T G C A T C C C A T C G A T C A T C G A T G T C A T 5040
C G T C G T C G A A A A G C A G C A A G C A G G A C A A G A T C A G C T T G A G C T C G T T T G G C A A G A A C A A G A A G A G C T G G A T 5110
C C G C T C C T C A C T C T C C A A G T T C A C C A A S A A G A A G A A C A A G A A C T A C G A C G A A G C A C A T A T G C C A T C A A T T 5180
T C C G G A T C T C A A G G A A C T C T T G A C A A C A T T G A T G T G A T T G A G T T G A A G C A A G A G C T C A A A G A A C G C G A T A 5250
6260 6270 6280 6290 6300 6310 6320
G T G C A C T T T A C G A A G T C C G C C T T G A C A A T C T G G A T C G T T C C C G C G A A G T T G A T G T T C T G A G G G A G A C A G T 5320
G A A T A A G T T G A A A A L L G A G A A C A A G C A A T T A A A G A A A G A A G T G G A C A A A C T C A C C A A C G G T C C A G C C A C T 5390
C G T G C T T C T T C C C C G C C C T C A A T T C C A S T T A T C T A C G A C G A T G A G C A T G T C T A T G A T G C A G C G T G T A G C A 5460
G T A C A T C A G C T A G T C A A T C T T C G A A A C S A T C C T C T G G C T G C A A C T C A A T C A A C G T T A C T G T A A A C G T G A 5530
C A T C G C T G G A G A A A T C A G T T C G A T C G T T A A C G G G A C T T G A A G C A G C A G G A A T T C T T C C T G G G C T G T A G C 5600
5610 5620 5630 5640 5650 5660 5670
A A G G T C A G T G G A A A A G T T G A C T G G A A G A T G C T G G A T G A A G C T G T T T T C C A A G T G T T C A A G G A C T A T A T T 5670
C T A A A A T G G A C C C A G C C T C T A C C C T C G S A C T A A G C A C T G A G T C C A T C C A T G G C T A C A G C A T C A G C C A C T 5740
G A A A C G A G T G T T G G A T G C A G A C C C C C C G A G A T G C C T C C T T G C C G T C G A G G T G T C A A T A A C A T A T C A G T C 5810
T C C C T C A A A G G T C T G A A G G A G A A A T G C C T C G A C A G C C T G G T G T T C G A G A C G C T G A T C C C C A A G C C G A T G A 5880
T C C A G C A C T A C A T A A G C C T C C T G C T G A A G C A C C G G C G C C T C C T C C T C T C G G C C C C A G C G G C A C G G G C A A 5950
5960 5970 5980 5990 6000 6010 6020
G A C C T A C C T G A C C A A T C G C T T G G C C G A S T A C C T G G T G G A G C G C T C T G G C C G T G A G S T C A C A G A G G G C A T C 6020
G T C A G C A C C T T C A A C A T G C A C C A G C A G T C T T G C A A G C A T C T G C A A C T G T A T C T T T C C A A C C T A G C C A A C C 6090
A G A T A G A C C G S G A A A C A G S A A T T G S G A T G T G C C C C T G S T G A T T C T A T T G S A T G A C C T G A G T G A A G C A G S 6160
C T C A T C A G T G A G T T G G T C A A T G G G S C C C T C A C C T G C A A G T A T C A T A A T G T C C C T A T A T T A T A G G T A C C 6230
A C A A T C A G C C T G T A A A A A T G A C A C C C A A C C A T G G C T T G C A C T T G A G C T C A G S A T G T T G A C C T T C T C A A 6300
6310 6320 6330 6340 6350 6360 6370
A C A A C G T G G A G C C A G C C A A T G G C T T C C T G S T T C G T T A C C T G A G G A G G A A G C T G G T A G A G T C A G A C A S C S A 6370
C A T C A A T G C C A A C A A G G A A G A G C T G C T T G S G S T G C T G A C T G S G T A C C C A A G C T G T G G T A T C A T C T C C A C 6440
A G C T T C C T T G A G A A S A G A G A G A G G T G G A S T T C C T C A T S C C C C T T C C T 6510
G C A T T G A G G A C T T C C S G A C C T G G T T C A T T G A C C T G T G G A A C A C T C A T C A T T C C C T A T C T A C A G G A A G S 6580
A S C C A A G G A T G G G A T A A A G G T C A A T G S A C A S A A A G C T G C T T G S G A G A C C C A G T G C A A T G G S T C C G G G A C 6650
6660 6670 6680 6690 6700 6710 6720
A C A C T T C C C T G G C C A T C A G C C C A A C A A G A C C A A T C A A A C C T G T A C C A C C T G C C C C A C C C A C C G T G G C C C 6720
C T C A C A G C A T T G C C T C A C C T C C C G A G G A T A G G A C A G T C A A A G A C A G C A C C C C A A G T T C T C T G S A C T C A G A 6790
T C C T C T G A T G C C A T T C T C T G A A A C T T C A A G A A G C T G C C A A C T A C A T T G A G T C T C C A G A T C A G A A A C C 6860
A T C C T G G A C C C C A A C C T T C A G G C A A C A C T T T A A G S S T T C G S C A A T C A C T G T C A C C C C C G G A C A C C A G A C 6930
G C T G C C A T C A G C T A T C T T A G C T C T C C T C T C C C C T C T C C T C T T C A G A G C A C T G G C T C T C A G C C C C A G C 7000
7010 7020 7030 7040 7050 7060 7070
A T G A G A A C A C G A G S G A G S A G G A G A T G A A S A G S A G G G A C A G G T T C T T G S T G C T G T A C C T T T G A G A A C T C 7070
C T A G C A A C S A A T G G T G S S G T G G C G T T G S G A A C T T G T G C C C C C T A A A C A C A T T A S T G G C C T C C T C A A T 7140
G A C T T T G G G G A A A G A T G A T T C T G G S T T T T C C C T T G A C T T C T T G T T C A A T T A C A A C T C C T G G C C T T T 7210
C T G G G A C C G G T T C A S A A A A T C A A A A C A C T G C A G C A G T T C C C C G G A A T T C A C C T T G A C T T A A A C C A S S 7280
C T G A A C T T G C T C A A A A G A A G C C G A A T T C A G C A A A C T G G C C T C C C C A T G G T A T T C A T A C T G A G C T C C C C 7350

Monday, 1 December 1997 14:12

Page 4

lg pCB501

7360 7370 7380 7390 7400 7410 7420
ATGGGCGGCGTGCATCAGATCGCCATCTCGCGCGCGTGCCTCTGACTTCFAAGGCCAATFACGCTTCAAC 7420
ATCCCTACATGCTCTTTCGCCGCGGCGCCACCCCTATTTTGTATTATCAAAAAAACCTCTCTTA 7430
ATTTCTTTGTTTTTACCTTCTTTAACTCAGCTCTAACAATGAAATTTGTGTAGATCAAAAAATAGATT 7440
AATTCGTAATAAAAAAGTCGAAAAAATTTGTCTCCCTCCCCCATTAATAATAATCTATCCCAAAATCT 7450
ACACAATGTTCTGTGTACACCTCTAAGTTTTTTTACTTCTGATAAATTTTTTTGAAACATCATAGAA 7460
7710 7720 7730 7740 7750 7760 7770
AAAACCGCACACAAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTTATTTCTTCGCA 7770
CGCTCGGCGCTCTCATGACGTCAAAATCATGCTCATCGTGA AAAAATTTTGGAGTATTTTGGAAATTTTC 7780
AATCAAGTGAAAGTTTATGAAATTAATTTTCTGCTTTTGTCTTTTGGGGGTTTCCCTATTGTTTGTCA 7790
AGAGTTTCGAGGACGGCGTTTTCTTCTTAAATCAACAAGTATTGATGAGCAGGATCAAGAAAGATCGG 7800
AAGAAGGTTTGGGTTGAGGCTCAGTGAAGGTGAGTGAAGTTGATAATTTGAAATGGAGTAGGCTCT 7810
8080 8090 8100 8110 8120
ATGGGGTTTTTGGCTTAAATGACAGAATACATTCCCAATATACCAACATAACTGTTTCTACTAGTGG 8120
CGTACGGGCGCTTTCTGCTCGCGGCTTTGGGTGATGACGGTGAAAACCTCTGACACATGACGCTCCCG 8130
AGACGGTCACAGCTTGTCTGTAAGCGGATGCGGGGAGCAGACAAGCCGTCAGGGCCTGTCAGCGGCTGT 8140
TGGCGGGTGTGCGGGCTGCTTAACTATGCGGCATCAGAGCAGATTGTAAGTGAAGAGTGCACCAATTCGG 8150
TGTGAAATACCGCACAGATGCGTAAAGGAAAAATACCGCAACAGGCGGCTTGAAGGCTCTGTAAGG 8160
8410 8420 8430 8440 8450 8460 8470
CTATTTTTATAGGTTAATGTCTATGATAATAATGGTTCTTACAGCTCAGGTGGCACTTTTCCGGGAAAA 8470
TGGCGGGAACCGCTATTTGTTTATTTTCTTAAATACATTCAAAATATGTAAGCTGTAAGAGACAAATACC 8480
CTGATAAATGCTTCAATAATATGAAAAAGGAAGATATGAGTATCAACATTTCTGTGTGCGCTTATT 8490
TCTTTTTTGGGGCATTTTGGCTTCTGTTTTTCTCACCAGAAAACGCTGGTGAAAGTAAAGATGCTG 8500
AAGATCAGTTGGGTCCAGGAGTGGGTTACAACGAAGTCTCAACAGCGCTAAGATCTTGAAGATTT 8510
8760 8770 8780 8790 8800 8810 8820
TGGCGCGGAAGAAGCTTTTCCAAATGATGAGCACTTTAAAGTTCTGCTATGTGCGGCGGTAATTAACCG 8820
ATTTGACGCGCGGCAAGACCAACTCGGTCCCGGATACACATTTTCAGAAATGACTTGGTTGAGTACTCAC 8830
GATTCACAGAAAAGCATCTTACGGATGCAATGACAGTAAGAGAATTATGCAATGCTGCCATTAACATGAG 8840
TGATAACACTCGCGGCAACTTACTTCTGACAAAGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC 8850
AAGATGGGGGATCATGTAACCTGCTTGGGTAACCGAGCTGAAATGAAGCCATACCAACAGGCT 8860
9110 9120 9130 9140 9150 9160 9170
AGGGTGACACCAAGATGCGCTGAGGCAATGGCAACAACGTTGGCGAAACTATTAACCTGGCGAACFAC 9170
CTAGCTTCCCGGCAACAATTAATGAGCTGGATCGAGCGGATAAAGTTGCAGGACCACTTCTGCGCTCG 9180
CGCTTCCGCTGCGTGGTTTATTGCTGATAAATCGGAGCGGTTGAGCGTGGCTCTCGCGGTAATTAATG 9190
TACCACTGGGGCAGATGGTAAGCCCTCTGCTATGATGATCTACACGACGGGGAGTCAGGCAACTAT 9200
AGATGAACGAATAGACAGATGCTGAGATAGGTGCTGATGATTAAGCATGGTAAGTCTGAGAGCA 9210
9460 9470 9480 9490 9500 9510 9520
GTTTACGATATATACTTTAGATGATTAAAACTTCAATTTTAAATTTAAAGGATCTAGGTAAGATTC 9520
TTTTTGAATCTCATGACCAAAAATGCTTAACTGAGTTTCTGTTCCACTGAGCGTCAGACCCCTAGAG 9530
AAGATCAAGGATCTCTTGAGATGCTTTTCTGCGGTAATCTGCTGCTTGAAGCAAAAAACCA 9540
CGGCTACCAAGGTTGTTGTTTCCCGGATCAAGAGCTACCAACTTTTTTCCGAAGGTAATGCTTCA 9550
GAGAGGCGAGATACCAATACTGCTCTCTAGTGTAGCTGAGTGAAGGCAACCTCAAGAACCTGCT 9560

Monday, 1 December 1997 14:12

Page 5

lig pCB501

```

      9810      9820      9830      9840      9850      9860      9870
-----
ASCACCGCCTACATACC TCGCTCTGCTAATCCTGTTACCACTGGCTGCTGCCAGTGTCGATAAAGTCGTGT 9870
CTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGGCTGAACGGGGGCTTCGCT 9940
GCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT3AGCATTGAGAAAG 10010
CGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCG3AACAGGAGAGCGC 10080
ACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTG 10150
      10160      10170      10180      10190      10200      10210      10220
-----
AGCGTCGATTTTTGTGATGCTCGTCA3GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGGCGGCTTTTT 10220
ACGGTTCCTTGGCCCTTTTCTGCTCCTTTTGCTCACA TGTTCCTTTCCTGCGTTATCCCCTGATTCTGTGGAT 10290
AACCGTATTACCGCTTTT3AGT3AGCTGATACCGCTCGCCGCGAGCCGAACGACCGA3CGCAGCGAGTCAG 10360
T3AGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCGCTCTCCCGCGCGTTGGCCGATTCAATTAATG 10430
CAGCTGGCACGACAGGTTTCCCGACT3GAAAGCGGGCAGT3AGCGCAACGCAATTAATGTGAGTTAGCTC 10500
      10510      10520      10530      10540      10550      10560      10570
-----
ACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAA TGT3AGCGGAT 10570
AACAA TTTACACAGGAAACAGCT 10594

```

SEQUENCE ID NO 10

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence

Enzymes : 100 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page

fig 13 pCB201

GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCC TAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGC CGGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTCATGAAGAATCTGCTTAGGGTTAGGCGTTTGGG
CCTCCAGCGACTCATCAGCGCTCGTTTAAATTCGATGTTGTTCCGTTCCGAACGGCTGTTAAGTACTTCTTAGACGAATCCCAATCCGCAAAACGG
G G R . V V R E Q N L S Y N K A R L D R Q L H E E S A . G . A F C

CTGCTTCGGCATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTATAGCCCATATA
GACGAAGCGCTACATGCCCGTCTATATGCCGAACGTGAACATAAAGTATCAATAATTATCATTAGTTAATGCCCGAGTAATCAAGTATCGGGTATA
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCCGCTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGCGGGTGTCTGGGGGCGGGTAACGACGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGGACTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCC
TTGGCGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAAATGCCATTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTTCATGCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTTACTTGGCAGTACATCTAGTATTAGTCA
GGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAG
P Y . R Q . R . N A R L A L C P V H D L M G L S Y L A V H L R I S H

TGCTATTACCATGGTGATSCGTTTTGGCAGTACATCAATGGCGTGGATAGCGGTTTGACTCAGGGGATTCCAAGTCTCCACCCCATGACGTCAA
AGCGATAATGGTACCCTACGCCAAACCGTCATGATGTTACCGGCACCTATCGCCAAACGAGTGCCCTAAAGGTTGAGAGGTGGGGTAAGTGCAGT
R Y Y H G D A V L A V H Q V A V I A V . L T G I S K S P P H . R Q

TGGAGTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAG
ACCTCAAAACAAACCGTGSTTTTASTTSCCCTSAAGGTTTTACAGCATTGTTGAGGCGGGTAACGCTTTTACCCGTCATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCAAGCTGGCTAGG
CAGATATATTGCTCTCGAGAGACGATTGATCTCTTGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGAGCCGATCG
G L Y K Q S S L A N . R T H C L L A Y R H . Y D S L . G D P S V L A

GTTTAAACTTAAGCTTACCATTGGGSGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTSSACAGCAAAATGGGTCGGGATCTGTACGAG
CAGATTTGAATCGAATGGTACCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATGCTACTGACCACTGCTGCTTTACCCAGCCCTAGACATGCTG
F < L K L T M G S S H - H H H H G M A S M T S G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GATGACGATAAGGTACCTAGGATCCATATGCCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACA
CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACASTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCCCTTTACGCAGCTG
pCB201 insert = U4
U4 ORF
D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C V D
GCCGCTGCTTTCGAGACGCTGATCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCAGCGCCTCGTCTCTCGGGCCCCAGCGGCA
CGGACCACAAGCTCTGCGACTAGGGGTTCCGCTACTACGTCGTGATGATTTCGGAGGACGACTTCGTGGCCCGCGGAGCAGGAGAGCCCGGGTCCGCGT
pCB201 insert = U4
U4 ORF
S L V F E T L I P K P M M O H Y I S L L L K H R R L V L S G P S G T
GGGCAAGACCTACCTGACCAATCGCTTGCCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCA
CCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGT
pCB201 insert = U4
U4 ORF
G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q
CACTCTTGAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTTGGTGATTCATTGSA
GTCAAGACGTTCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCAC TAASATAACCTAC
pCB201 insert = U4
U4 ORF
V S C K D L O L Y L S N L A N O I D R E T G I G D V P L V I L L D
ACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCAACATCAGCTG
TGGACTCAGCTTCTGTCGAGGTAGTCACTCAACCACTTACCCCGGAGTGGAGGTTCAATGATTACAGGGATATAATATCCATGGTGGTATGTCGAC
pCB201 insert = U4
U4 ORF
D L S E A G S I S E L V A G A L T C K Y H K C P Y I I G T T N Q F
AAAAATGACACCCAACCATGGCTTGCCTTGAAGTTCAAGATGTTGACCTTCTCAACCAACGTGGAGCCAGCCAATGGCTTCTGGTTGGTTACCTGAG
TTTCTACTGTGGGTGGTACCGAAGCTGAAGTCAAGTCTTCAACCAACGTGGAGGTTGTTGACCTCGGTGGTTACCGAAGGACCAAGCAATGGAGTCT
pCB201 insert = U4
U4 ORF
K N T P N H G L H L S F R M L T F S N H V E P A N G F L V R Y L F

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAAGAGCTGCTCGGGTGCCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACC
TCCCTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTCTCGACGAAGCCACGAGCTGACCCATGGGTTGACACCATAGTAGAGGTGTGGG

pCB201 insert = U4

U4 ORF

R K L V E S D S D I N A N K E E L L R V L D V V P K L V Y H L H T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTTCTTGTGCTGTCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCT
AGGAACCTCTTCGTGCTGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCCGACCAAGTAACGGG

pCB201 insert = U4

U4 ORF

F L E K H S T S D F L I G P C F F L S C P I G I E D F R T V F I D L

GTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGT
CACCTTGTGAGATAGTAAGGGATAGATGTCTTCTCGGTTCTTACCCTATTTCCAGGTACCTGTCTTTCGACGAACCTCTCTGGGTCACCTTACCCAG

pCB201 insert = U4

U4 ORF

V N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D P V E W V

CGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCCTACCTCCCG
GCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTCTGGTTAGTTTCGACATGCTGGACGGGGGTGGGTGGCACCCGGGAGTGTCTGAACGGAGTGGAGGGG

pCB201 insert = U4

U4 ORF

R D T L P V P S A Q Q D Q S K L Y H L P P P T V G P H S I A S P P

AGGATAGGACAGTCAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAATACATAGAGTC
TCCATCTCTGTCAGTTTCTGCTGTGGGGTTCAAGAGACC TGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTTCTTCGACGTTGATGTAACTGAC

pCB201 insert = U4

U4 ORF

E D R T V K D S T P S S L D S D P L M A M L L K L Q E A A H Y I E S

TCCAGATGGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTAAAGGTTCTGGCAATCACTGTACCCCCGGACAGCAGAACGCTGGCATCAGCTA
AGGTCATCTCTTTGGTAGGACCTTGGGTTGGAAGTCCGTTGTGAAATTCCEAAGCCGTTAGTGACAGTGGGGGCCCTGTCGCTTTGCGACCGGTAGTGG

pCB201 insert = U4

U4 ORF

P D F E T I L C P N L Q A T L G F G N H C H P R T A E R V H Q L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

TCTTAGCTCCCTCCTCTCCCTCTCTTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGAGATGAAAAGGAGGGACAGGT
AGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCAGAGAGTGGGGTCTCTCTTGTCCTCCCTCTCTACTTTCTCTCTCTGTCCAA

pCB201 insert = U4

S L L L S P L L F Q S T G S P A P G G E Q E G G G O E R G G T G

CTTGGTGCTGTACCTTTGAGAATTCTCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTGTGCCCTTAACACATTTACTGGCTCTCTAATGACT
GAACCACGACATGGAACCTCTTGAAGGATCCTCTTACCACCCCACCGCAAACCTTGAACACGGGGGATTGTGTAAATGACCGAGGAGATTACTGA

pCB201 insert = U4

S V C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TTGGGGAAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTITTAATTACAACCTCTGGGCTTCTGGGGAGGGGTTACAGAAAACATCAAAACACTGC
AACCCCTTTTCTACTAAGACCCAGAAAGGGAACGAAGAACAAGTTAATGTTTGAGGACCCGAAAGACCCCTCCCAAGTCTTTGTAGTTTTGTGAGC

pCB201 insert = U4

V G K D D S G S F P . L L V S I T N S V A F V G G V Q K T S K H C

AGCAGTTCTAAATGATTCTCACAAAGCAACCTGAGAGAGACAGTCTTGTGAGGGAGATCTGGGGGAGGAGGAAAGCTCTCAGATTTTCTCACAGACC
TCGTCAAGGATTTACTAAGAGTGTTCTGTGGGACTCTCTCTGTGAGAACACTCCCTCTAGACCCCTCCGTCTTCGAGGAGTCTAAAGAGTGCTGGG

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I V G R Q E A P O I F S Q T

TTCCAATTCCATCACCCTGCAACACTCGTCEGGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTG
AAGGTTAAGGTAGTGGTGACGGTTGTGAGCAGGCTTAAGACGCTATAAGTCTGTGTCACCGCCGGGAGCTCAGATCTCCCGGSCAAATTTGGGCGAC

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q V R P L E S R G P V . T R .

ATCAGCTCGACTGTGCCTTCTAGTTGCCAGCATCTGTGTTTGCCCTTCCCCGTGCCCTTCTTGACCTTGAAGGTGCCACTCCCACTGTCTTTCC
TAGTCGGAGCTGACACGGAAGATCAACGGTTCGGTAGACAACAACCGGGGAGGGGACGGAAGGAACCTGGGACCTTCCACGGTGAGGGTGACAGGAAAGG
S A S T V P S S C O P S V V C P S P V P S L T L E G A T P T V L S

TAATAAAATGAGGAAATGTCATCGCATTTGTCTGAGTAGGTGTCTATTCTTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATTATTTTACTCCTTTAAGGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCACCCCAACCCGTCTCTGCTTCCCCCTCTTAACCCCTCTGT
. . N E E I A S H C L S R C H S I L G G G V G Q D S K G E D W E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGGTATCCCACCGGCCCTTAGCGGGC
TATCGTCCGTACGACCCCTACGCCACCCGAGATACCAAGACTCCGCTTTCTTGGTCGACCCGAGATCCCCATAGGGGTGCGCGGGACATCGCCGGS
N S R H A G D A V G S M A S E A E R T S W G S R G Y P H A P C S G A

ATTAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTACCGCTACACTTGCAGCGCCCTACGCGCCGCTCTCTTCGCTTTCTTCCCTTCTTTCTCGCT
TAATTCGCGCGCCCAACACCAATGCGCGTCSACTGCGGATGTGAACGCTCGCGGGATCGCGGGCAGGAAAGCGAAAGAGGAAAGGAGAGG
L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACGTTCCCGGGCTTTCCCGCTCAAGCTCTAAATCGGGGATCCCTTTAGGGTTCGATTTAGTGCTTTACGGCACCTCGACCCCCAAAAAATTGATTAGG
TGAAGCGCGCGGAAAGGGGAGTTGAGATTTASCCCGTAGGGAAATCCCAAGGCTAAATCACGAATGCCGTGGAGCTGGGGTTTTGTAACTAATCT
T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GTGATGGTTACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCACAACTGG 330
CAC TACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAGTCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAAGGTTTGACC
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G
AACAACACTCAACCTATCTCGGTCTATTCTTTGATTTATAAGGGATTTTGGGGATTTTCGCCCTATTGGTTAAAAATGAGCTGATTTAACAAAAATTT 340
TTGTTGTAGTGGGATAGAGCCAGATAAGAAAACATAATATTCCTAAAACCCCTAAAGCGGGATAACCAATTTTACTCGACTAAATTTGTTTTAA
T T L N P I S V Y S F D L . G I L G I S A Y W L K N E L I . Q K F
AACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCGAGGCTCCCGAGGCAGGAGAGTATGCAAGCATGCATCTCAATTAGT 350
TTGCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTTCAGGGGTCGAGGGGTCGCTCGTCTTCATACGTTTCGTACGTAGAGTTAATCA
N A N . F C G M C V S . G V E S P Q A P Q A G R S M Q S M H L N .
CAGCAACCGAGGTGGGAAAGTCCCGAGGCTCCCGAGCAGGAGAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCC 360
GTCGTTGGTCCACACCTTTTCAGGGGTCGAGGGGTCGCTCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGCGGGGATTGAGG
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P
GCCATCCCGCCCTAACTCCGCCAGTTCGCCCATTCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCT 370
CGGGTAGGGCGGGATTGAGGCGGGTCAAGGCGGGTAAGAGCGGGGTAACGACTGATTAATAAATAAATACGCTCCGGCTCCGGCGGAGACGGAGG
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S
GAGCTATCCAGAGTAGTGAGGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGA 380
CTCGATAAGGCTTCATCACTCTCCGAAAAACCTCCGGATCCGAAAAAGCTTTTCGAGGGCCCTCGAACATATAGGTAAGGCTAGAGTAGTTCTCT
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E
CAGGATGAGGATCGTTTCGCATGATTGAACAAGAAGGATTCGACGCGAGGTTCTCCGGCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACA 390
GTCTTACCTCAGCAAGCGTACTAATCTGTTCTACCTAACGTGCGTCCAAGAGGCGCGGCAACCCACCTCTCCGATAAGCCGATCTGACCCGTGTTG
T S . G S F R M I E Q D G L H A G S P A A W V E R L F G Y D W A Q Q
GACAATCGGCTGCTCTGATGCCCGGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAAGT 400
CTGTTAGCGAGGAGACTACGGCGGCACAGGCGGACAGTCCGCTCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCGCAGGGACTTACTTGAC
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L
CAGGACGAGGAGCGGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGGCAGCTGTGCTCGACGTTGTCTACTGAAGCGGGAAGGAGCTGGCTGCTA 410
GTCTGCTCCGTCGCGCGGATAGCACCGACCGGCTGCTGCCCGCAAGGAACGCTGACACGAGCTGCAACAGTACTTCGCCCTTCCTGACCGAGGATA
S D E A A R L S V L A T T G V P C A A V L D V V T E A G R D V L L
TGGCGAAGTGCCGCGGAGGATCTCTGTCATCTCACCTTGTCTCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCAATAGCTTGG 420
ACCGCTTCACGCGCCCGTCTAGAGGACAGTASAGTGGAAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCCGGAGCTATGCGAAT
L S E V P S Q D L L S S H L A P A E K V S I M A D A H R R L H T L D
TCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAA 430
AGCGGATGAGGAGGTAAGCTGGTGGTTTCGCTTTGAGCGTAGCTCGCTCGTGCATGAGGCTACCTTCGGCCAGAACAGCTAGTCTCTACTAGAGCTGCTT
P A T C P F D H O A K H R I E R A R T R M E A G L V D Q D D L D E
GAGCATCAGGGCTCTCGCCAGCGCAACTGTTCTCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCTGCTTGG 440
CTGCTAGTCCCGAGCGGCTGGCTTGACAGGCGGTCGAGTTCCGCGCTACGGGCTGCCGCTCTAGAGCAGCACTGGGTACCGCTACGGAGCGAGG
E H G G L A P A E L F A R L K A R M P D G E D - L V V T H G D A C L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

CGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTTACCGGCGAAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCCTGGCGATAGTCCGTATCGCAACCGATGGGCACT 500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D
TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGTGCTTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACCGCCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 560
I A E E L G G E V A D R F L V L Y G I A A P D S Q R I A F Y R L L
GACGAGTTCCTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 570
D E F F . A G L V G S K . P T K R R P T C H H E I S I P P P P S M
AAGGTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCCTCGCCACCCCAACTTGTTTATT
TTCCAACCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 580
K G V A S E S F S G T P A G . S S S A G I S C V S S S S P T P T C L L
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGCCAAACTCATCA
CGTCGAATATTACCAATGTTTATTTTCGTTATCGTAGTGTAAAGTGTATTTTCGTAAGGAGTGAAGTCAACACCAACAGGTTTGAGTAGT 590
Q L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S
ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATGTTATCCGCTCACAAT
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTA 500
M Y L I M S V Y R R P L A R A V R N H G H S C F L C E I V I R S Q
TCCACACAACATACGAGCGGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG 5082
AGGTGTGTTGTATGCTCGGCCCTTCGATTTTACATTTTCGGACCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAAC
F H T T Y E P E A . S V K P G V P N E . A N S H . L R V

Claims

1. A vertebrate protein homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative or bioprecursor thereof, which protein comprises an amino acid sequence having a statistically significant homology to the amino acid sequence of said UNC-53 protein of C. elegans illustrated in Figure 2.
2. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D or E as illustrated in Figure 9a, or block F in Figure 12a or a sequence having a statistically significant homology therewith.
3. A vertebrate protein homologue of an UNC-53 protein of C. elegans, which protein comprises an amino acid sequence having one or more of sequence blocks A, B, C, D, E or F which differ from those blocks of Figure 9a or 12a only in conservative amino acid changes.
4. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown from nucleotide positions 1 to 6013 illustrated in Sequence ID No. 3.
5. A vertebrate protein comprising an amino acid sequence which comprises one or more of the prosite signatures as illustrated in Figure 28 for each of said sequences of homology as claimed in claim 2.
6. A vertebrate protein comprising an amino acid

- 185 -

sequence as claimed in any one of claims 1 to 6 which is a human protein or a mouse protein.

5 7. A vertebrate protein having an amino acid sequence encoded by the nucleotide sequence shown in Sequence ID No. 4.

10 8. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 in one or more conservative amino acid changes.

15 9. A vertebrate protein homologue according to any one of claims 1 to 7 comprising an amino acid sequence as shown in Sequence ID No. 2 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 in one or more
20 conservative amino acid changes.

25 10. A cDNA encoding a vertebrate homologue of UNC-53 protein of C. elegans according to any of claims 1 to 9.

30 11. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 1 or an amino acid sequence which differs from the amino acid sequence shown in Sequence ID No. 1 only in one or more conservative amino acid changes.

35 12. A cDNA according to claim 10 comprising a sequence of nucleotides encoding an amino acid sequence as shown in Sequence ID No. 2 or an amino

- 186 -

acid sequence which differs from the amino acid sequence shown in Sequence ID No. 2 only in one or more conservative amino acid changes.

5 13. A cDNA according to any of claims 10 or 11 which cDNA comprises at least from nucleotide position 1 to position 6013 of the sequence as shown in Sequence ID No. 3.

10 14. A cDNA according to claim 10 or 12 which comprises the nucleotide sequence illustrated in Sequence ID No. 4.

15 15. A nucleic acid molecule capable of hybridising to the DNA sequences according to any of claims 10 to 14 under high stringency conditions.

20 16. A DNA expression vector which comprises a cDNA as claimed any of claims 10 to 14.

25 17. A vector according to claim 16 which comprises a promoter of C. elegans UNC-53 protein or a vertebrate homologue thereof according to any of claims 1 to 9.

30 18. A vector according to claim 17 wherein said promoter sequence is derived from a gene encoding a mouse or human homologue of an UNC-53 protein of C. elegans.

35 19. A vector according to any of claims 16 to 18 which further comprises a sequence encoding a reporter molecule.

20. A vector according to claim 19 wherein said

- 187 -

reporter molecule is a fluorophore.

21. A host cell transformed or transfected with the vector of any of claims 16 to 20.

5

22. A host cell transformed or transfected with the vector of claims 19 or 20.

23. A host cell according to claims 21 or 22,
10 which cell comprises a prokaryotic cell such as a bacterial cell or a eukaryotic cell such as a fungal, an animal, a plant or an insect cell.

24. A transgenic cell, tissue or organism
15 comprising a transgene capable of expressing a protein according to any of claims 1 to 9.

25. A transgenic cell, tissue or organism according to claim 24 which comprises any of a COS
20 cell, Hep G2, MCF-7 cell, N4 mouse neuroblastoma cell, a NIH3T3 cell, or colorectal carcinoma or human derived cells.

26. A transgenic cell, tissue or organism
25 according to claim 24 or 25 wherein said transgene comprises a vector according to any of claims 16 to 20.

27. A transgenic cell, tissue or organism
30 according to claim 24 to 26 wherein said transgene comprises a vector according to claim 19 or 20.

28. A transgenic cell, tissue or organism according to any of claims 24 to 26 wherein said
35 organism comprises any of an insect, a fungus, a non-

- 188 -

human mammal, a plant or a nematode worm.

29. A method of producing a mutant vertebrate non-human organism which mutation affects cell
5 behaviour or the regulation of cell motility or the shape or the direction of cell migration, which method comprises inducing a mutation in the wild type gene encoding the vertebrate homologue of an UNC-53
10 C. elegans protein.

30. A vertebrate protein homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, for use
15 as a medicament to promote neuronal regeneration, revascularisation, wound healing or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

31. A vertebrate protein homologue of an UNC-53
20 protein of C. elegans for use as claimed in claim 30 wherein said vertebrate human homologue is as claimed in any one of claims 1 to 9.

32. Use of a vertebrate protein homologue of an
25 UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor thereof, in the manufacture of a medicament for promoting neuronal regeneration, revascularisation, wound healing or for treatment of chronic
30 neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

33. Use of a vertebrate protein homologue of
35 UNC-53 protein of C. elegans according to claim 32 wherein said vertebrate protein homologue is as

- 189 -

claimed in any one of claims 1 to 9.

34. A pharmaceutical composition comprising a vertebrate homologue of an UNC-53 protein of C. elegans, or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate protein, together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

35. A pharmaceutical composition as claimed in claim 34 which comprises a vertebrate homologue of an UNC-53 protein of C. elegans according to any of claims 1 to 9.

36. A nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, for use as a medicament.

37. A nucleic acid sequence according to claim 36 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said cDNA sequence.

38. Use of a nucleic acid sequence encoding a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue, in the manufacture of a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

- 190 -

39. Use of a nucleic acid sequence according to claim 38 wherein said sequence is a cDNA sequence as claimed in any of claims 10 to 14 or a functional fragment of said nucleic acid sequence.

5

40. A pharmaceutical composition comprising a nucleic acid sequence according to claim 36 or 37 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

10

41. A pharmaceutical composition according to claim 40 wherein said nucleic acid sequence is a cDNA sequence as claimed in any of claims 10 to 14.

15

42. A method of determining whether a compound is an inhibitor or enhancer of the regulation of cell behaviour, growth, cell shape or motility or the direction of cell migration, which method comprises contacting said compound with a host cell according to claim 21 or 23 or a transgenic cell as claimed in any of claims 24 to 27 and screening for a phenotypic change in said cell.

20

43. A method according to claim 41 which is capable of determining whether said compound is an inhibitor or an enhancer of the signal transduction pathway of said transgenic cell of which said vertebrate homologue of an UNC-53 protein or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue is a component or is an inhibitor or an enhancer of a parallel or redundant signal transduction pathway in said cell.

30

44. A method according to claim 43 wherein said

35

- 191 -

method is capable of determining whether said compound is an inhibitor or an enhancer of said vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

45. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in cell growth, or shape or a change in cell motility.

46. A method according to any of claims 42 to 44 wherein said phenotypic change to be screened is a change in filopodia outgrowth, ruffling behaviour, cell adhesion, contact inhibition or the length of neurite growth.

47. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an N4 neuroblastoma cell and the phenotypic change to be screened is the length of neurite growth.

48. A method as claimed in any of claims 42 to 44 wherein said transgenic cell is an MCF-7 breast carcinoma cell or an NIH3T3 cell and the phenotypic change to be screened is the extent of phagokinesis or contact inhibition.

49. A method of determining whether a compound is an inhibitor or an enhancer of the regulation of cell shape, cell growth or motility or of the direction of cell migration, which method comprises administering said compound to a transgenic organism according to any of claims 24 to 28 or a mutant organism produced according to the method of claim 29

- 192 -

and screening for a phenotypic change in said organism.

50. A method according to claim 49, wherein said
5 method is capable of determining whether said compound
is an inhibitor or enhancer of a protein of the signal
transduction pathway of said transgenic or mutant
organisms, of which the vertebrate homologue of UNC-53
protein of C. elegans or a functional equivalent,
10 derivative, fragment or bioprecursor of said
vertebrate homologue is a component, or is an
inhibitor or an enhancer of a parallel or redundant
signal transduction pathway in said cell.

51. A method according to claim 50 wherein said
15 method is capable of determining whether said compound
is an inhibitor or an enhancer of the vertebrate
homologue of UNC-53 protein itself or a functional
equivalent, fragment, derivative or bioprecursor of
20 said vertebrate homologue.

52. A compound which is identifiable by the
method according to any one of the claims 42 to 51 as
an enhancer of the regulation of cell shape, or growth
25 or motility or the direction of cell migration for use
as a medicament for promoting neuronal regeneration,
revascularisation or wound healing or for treatment of
chronic neurodegenerative diseases or acute traumatic
injuries or fibrotic disease.

53. Use of a compound which is identifiable by
30 the method according to any one of the claims 42 to 51
as an enhancer of the regulation of cell shape, or
growth or motility or the direction of cell migration
35 in the preparation of medicament for promoting

- 193 -

neuronal regeneration, revascularisation or wound healing or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

5

54. A pharmaceutical composition comprising a compound identified according to the method of any of claims 42 to 51 claim and a pharmaceutically acceptable carrier, diluent or excipient therefor.

10

55. A compound which is identifiable by the method according to any one of claims 42 to 51 as an inhibitor of the regulation of cell motility, growth, or shape, or the direction of cell migration, for use as a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

15

56. Use of a compound according to claim 55 in the manufacture of a medicament for alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

20

57. A pharmaceutical composition comprising the compound as claimed in claim 55, and a pharmaceutically acceptable carrier diluent or excipient therefor.

25

58. A method of determining whether a compound is an inhibitor or an enhancer of transcription of a gene encoding a vertebrate homologue of UNC-53 protein of C. elegans, which method comprises the steps of (a) contacting said compound with a cell according to any of claims 22 or 27 and (b) monitoring the level of said reporter molecule and comparing the results

30

35

- 194 -

obtained from said monitoring step with a control comprising a cell according to claims 22 or 27, which cell has not been contacted with said compound.

5 59. A method as claimed in claim 58 wherein said reporter molecule detected is mRNA or green fluorescent protein.

10 60. A compound which is identifiable by the method according to claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, for use in promoting neuronal regeneration, revascularisation or wound
15 healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

20 61. Use of a compound which is identifiable by the method of claims 58 or 59, as an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for promoting neuronal regeneration,
25 revascularisation or wound healing, or for treatment of chronic neuro-degenerative diseases or acute traumatic injuries or fibrotic disease.

30 62. A pharmaceutical composition which comprises the compound of claim 60 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

35 63. A compound which is identifiable by the method of claims 58 or 59 as an inhibitor of transcription of a gene coding for a vertebrate

- 195 -

homologue of a UNC-53 protein of C. elegans or a functional fragment of said gene for use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

5

64. Use of a compound which is identifiable by the method of claims 58 or 59 as an inhibitor of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene, in the manufacture of a medicament for alleviating spread of disease inducing cells or metastasis or loss of contact inhibition.

65. A pharmaceutical composition which comprises the compound of claim 63 and a pharmaceutically acceptable carrier, diluent or excipient therefor.

66. A kit for determining whether a compound is an enhancer or an inhibitor of the regulation of cell motility, growth or shape or the direction of cell migration which kit comprises at least one transgenic cell as claimed in any one of claims 22 to 25 to be contacted with said compound and at least one cell according to claims 21 to 28 to be used as a control and means for contacting said compound with one of said at least one transgenic cells.

67. A kit for determining whether a compound is an inhibitor or an enhancer of transcription of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional fragment of said gene which kit comprises at least one cell as claimed in any one of claims 21 to 25 means for contacting said compound with said cells.

- 196 -

68. A kit for determining whether a compound is an enhancer or an inhibitor of the activity of a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor of said vertebrate homologue protein, which kit comprises at least, one vertebrate mutant non-human organism produced according to the method as claimed in claim 29 or a transgenic organism as claimed in claims 24 to 28 and a wild type of said vertebrate mutant organism.

69. A method of identifying vertebrate homologues of an unc-53 gene of C. elegans or a functional fragment thereof, which method comprises hybridizing to a DNA library a suitable oligonucleotide sequence of between 15 to 50 nucleotides of the nucleic acid sequence encoding unc-53 or a functional equivalent, derivative or bioprecursor thereof, under appropriate conditions of stringency to identify genes having statistically significant homology with the cDNA according to any of claims 10 to 14.

70. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said vertebrate homologue is a component, which method comprises:

- (a) contacting an extract of said cell with an antibody to the vertebrate homologue of the UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said protein,
- (b) identifying the antibody/vertebrate

- 197 -

homologue complex, and

(c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein of C. elegans other than the antibody.

5

71. A method of identifying a further protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein or a functional equivalent, fragment or bioprecursor of said UNC-53 protein is a component, which method comprises:

10

(a) forming an antibody to the first identified protein bound to the vertebrate homologue of UNC-53 protein of C. elegans in claim 70,

15

(b) contacting a cell extract with said antibody and identifying the antibody/protein complex,

20

(c) analysing the complex to identify any further protein bound to the first protein other than the antibody, and

(d) optionally repeating steps (a) to (c) to identify further proteins in said pathway.

25

72. A method of identifying a protein which is active in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment or bioprecursor of said homologue protein is a component, which method comprises

30

(a) contacting an extract of said cell with the vertebrate homologue of an UNC-53 protein of C. elegans or a functional

35

- 198 -

equivalent, derivative or bioprecursor of said vertebrate homologue,

(b) identifying any vertebrate homologue of UNC-53 protein/protein complex formed and

5 (c) analysing the complex to identify any protein bound to the vertebrate homologue of UNC-53 protein other than the same vertebrate homologue of UNC-53 protein.

10 73. A method according to claim 72 which further comprises contacting a cell extract with any protein identified from step (c) not being the same as the vertebrate homologue of UNC-53 protein used and repeating steps (b) and (c) so as to identify any
15 further protein involved in the signal transduction pathway of said cell.

20 74. A method of identifying a protein involved in the signal transduction pathway of a cell of which a vertebrate homologue of an UNC-53 protein of C. elegans is a component which method comprises:

(a) providing an appropriate host cell having a DNA construct comprising a reporter gene under the control of a promoter
25 regulated by a transcription factor having a DNA binding domain and an activating domain,
(b) expressing in said host cell a first hybrid DNA sequence encoding a first fusion
30 of a fragment or all of a DNA sequence according to any of claims 10 to 14 and either said DNA binding domain or the activating domain of the transcription factor,
(c) expressing in the host cell at least
35 one second hybrid DNA sequence encoding a

- 199 -

putative binding protein to be investigated together with the DNA binding or activating domain of the transcription factor which is not incorporated in the first fusion,

5 (d) detecting any binding of the protein being investigated with a protein according to any of claims 1 to 9 by detecting for the production of any reporter gene product in said host cell.

10

75. A protein identified by the method of any one of claims 70 to 74 for use as a medicament to promote neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neuro-
15 degenerative diseases or acute traumatic injuries or fibrotic disease.

76. Use of a protein identified by the methods of any one of claims 70 to 74 in the manufacture of a
20 medicament for promoting neuronal regeneration, revascularisation or wound healing, or for treatment of chronic neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

25 77. A pharmaceutical composition comprising a protein identified by the methods of any one of claims 70 to 74 and a pharmaceutically acceptable carrier, diluent, or excipient therefor.

30 78. A process for producing a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent fragment, derivative or bioprecursor of said vertebrate homologue which process comprises culturing the cells of any of claims
35 21 to 28 and recovering said vertebrate homologue of

- 200 -

UNC-53 protein expressed.

79. A process for producing a vertebrate
homologue of an UNC-53 protein of C. elegans or a
5 functional equivalent, fragment, derivative or
bioprecursor of said protein which process comprises
culturing an insect cell transfected with a
recombinant Baculovirus vector, said vector comprising
a DNA insert encoding said vertebrate homologue of
10 UNC-53 protein or a functional equivalent, fragment or
bioprecursor of said vertebrate homologue, downstream
of the Baculovirus polyhedrin promoter, and recovering
the expressed vertebrate homologue of UNC-53 protein.

15 80. A nucleotide sequence comprising the
sequence as shown in figure 15.

81. A nucleotide sequence comprising the
sequence as shown in figure 16.

20 82. A nucleotide sequence comprising the
sequence as shown in figure 17.

83. A method of detecting whether a compound is
25 an inhibitor or an enhancer of expression of a
vertebrate homologue of an UNC-53 of C. elegans, or a
functional equivalent, derivative or fragment of said
vertebrate homologue which method comprises contacting
a cell expressing said homologue with said compound
30 and monitoring for a phenotypic change compared to a
control cell which has not been contacted with said
compound.

84. A method according to claim 83 wherein said
35 cell comprises a cell according to any of claims 21 to

- 201 -

28.

85. A method according to claim 83 wherein said cell has undergone loss of contact inhibition.

5

86. A method according to any of claims 83 to 85 which is capable of determining whether said compound is an inhibitor of expression of said vertebrate homologue in which the compound to be tested comprises a nucleic acid.

10

87. A method according to claim 86 wherein said nucleic acid sequence comprises an antisense DNA or RNA sequence.

15

88. A method according to claim 87 wherein said mRNA sequence comprises 3' untranslated regions of mRNA encoding for said vertebrate homologue.

20

89. A method according to any of claims 83 to 85 wherein said compound to be tested comprises a protein having an amino acid sequence potentially suitable for inhibiting function of said vertebrate homologue.

25

90. A method according to claim 89 wherein said protein comprises a protein identified according to any of the methods of claims 70 to 74.

30

91. A pharmaceutical composition comprising a compound identified according to any of claims 83 to 89 together with a pharmaceutically acceptable carrier, diluent or excipient therefor.

35

92. A nucleic acid sequence identified according to the method of any of claims 86 to 88 for use in

- 202 -

treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor thereof.

93. Use of a nucleotide sequence identified according to the method of any one of claims 86 to 88 in the preparation of a medicament for the treatment of loss of contact inhibition or carcinoma which is mediated by a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, fragment, derivative or bioprecursor of said vertebrate homologue.

94. A nucleic acid according to claim 92 for use in the preparation of a medicament for inhibiting expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans.

95. A NIH3T3 cell line transfected with pcB201 and deposited under LMBP Accession No. 1603CB.

96. A plasmid pcB 201 of Sequence ID No. 10 deposited under LMBP Accession No. LMBP 3594.

97. A MCF-7 cell line transfected with plasmid pcB 201 deposited under LMBP Accession No. LMBP 1601CB.

98. An assay for detecting expression of a vertebrate homologue of UNC-53 protein of C. elegans in a vertebrate cell which assay comprises contacting a cell or an extract thereof with an antibody to said vertebrate homologue, or a functional equivalent,

- 203 -

derivative or bioprecursor thereof, which antibody is linked to a reporter molecule, removing any unbound antibody and monitoring for the presence of said reporter molecule.

5

99. An assay according to claim 98 wherein said reporter molecule is an antibody conjugated with a suitable fluorophore or detectable enzyme.

10

100. A method for detecting for expression of a gene coding for a vertebrate homologue of an UNC-53 protein of C. elegans or a functional equivalent, derivative, fragment or bioprecursor thereof, which method comprises contacting a probe specific for a nucleic acid or protein sequence coding for or corresponding to said vertebrate homologue or a functional equivalent, fragment or bioprecursor therefor with a cell extract which probe is linked to a reporter and analysing for the presence of said reporter.

20

101. A method according to claim 100 wherein said probe comprises a complimentary sequence to a region of mRNA transcribed from said gene encoding said vertebrate homologue of UNC-53 protein or a functional equivalent, derivative or bioprecursor therefor.

25

102. A method according to claim 101 wherein said complimentary sequence is a 3' or 5' untranslated region of said mRNA.

30

103. A method according to claims 100 or 102 wherein said reporter comprises a radiolabel.

35

- 204 -

104. A method according to claim 100 wherein
said probe comprises an antibody specific for said
vertebrate homologue of said UNC-53 protein or a
functional equivalent, derivative, fragment or
5 bioprecursor therefor.

105. A method according to claim 104 wherein
said reporter comprises an antibody conjugated with a
detectable fluorophore or enzyme.
10

106. Phage Lambda clone 3b of Sequence ID No. 5
deposited under Accession No. LMBP 3595.

107. A method of determining whether a compound
15 is an inhibitor or an enhancer of association of UNC-
53 or a vertebrate homologue thereof according to any
of claims to 1 to 9 to microtubules or plus end
regions thereof, which method comprises:-

(a) contacting said compound with a
20 transgenic cell, tissue or organism
expressing UNC-53 protein or said vertebrate
homologue and which protein is operably
linked to a reporter molecule.

(b) screening for the localisation of said
25 reporter molecule as compared to a cell
according to step (a) which has not been
contacted with said compound.

108. A compound identifiable by the method
30 according to claim 107.

109. A compound identifiable by the method
according to claim 107 as an inhibitor of localisation
or association of UNC-53 or said vertebrate homologue
35 with microtubules or the plus end region thereof for

- 205 -

use in alleviating the spread of disease inducing cells or metastasis or loss of contact inhibition.

110. A compound identifiable by the method
5 according to claim 107 as an enhancer of association of UNC-53 or said vertebrate homologue with microtubules or the plus end region thereof, for use in promoting neuronal regeneration, revascularisation or wound healing, or for treating chronic
10 neurodegenerative diseases or acute traumatic injuries or fibrotic disease.

111. A pharmaceutical composition comprising the compound according to claims 108 or 109 and a
15 pharmaceutically acceptable carrier, diluent or excipient therefor.

112. A kit for determining whether a compound is an inhibitor or an enhancer of association of UNC-53
20 or a vertebrate homologue thereof according to any of claims 1 to 9 with microtubules or the plus end regions thereof, which kit comprises at least one transgenic cell expressing UNC-53 and a reporter molecule or a cell according to any of claims 20 to 24
25 and at least one cell of the same cell type for use as a control and means for contacting said compound with one of said at least one transgenic cells.

113. A composition comprising UNC-53 of C. elegans
30 elegans or a vertebrate homologue thereof according to any of claims 1 to 9 linked to a compound identified as an inhibitor or enhancer of association of UNC-53 or said vertebrate homologue with microtubules or their plus end regions for use in targeting said
35 compound to said microtubule or the plus end regions

- 206 -

thereof.

114. A composition according to claim 113 which further comprises a cell transformation or
5 transfecting agent.

115. A method of targeting a protein to a cell microtubule or the plus end region thereof, which method comprises introducing into a host cell, tissue
10 or organism a transgene comprising a sequence capable of expressing UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which sequence is operably linked to a sequence encoding said protein to be targeted such that a chimeric protein is expressed
15 and which results in targeting said protein to said microtubule or a plus end region thereof.

116. A method of identifying a molecule which covalently modifies UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which
20 method comprises

a) contacting either an extract from a cell expressing UNC-53 or said vertebrate homologue or a mixture of enzymes comprising candidate UNC-53
25 modifying enzymes in the presence of an indicator of covalent modification of a protein,

b) identifying any covalently modified UNC-53 protein from step a),

c) identifying said molecule involved in said
30 modification step.

117. A method according to claim 112, wherein said indicator comprises a p.

35

- 207 -

118. A method of identifying a compound which alleviates or enhances the toxicity of UNC-53 or a vertebrate homologue thereof according to any of claims 1 to 9, which method comprises contacting said compound with a cell, tissue or organism according to claim 27, and monitoring for the presence of said reporter molecule adjacent said microtubules or the plus end regions thereof.

119. Plasmid pLM1 of Sequence ID No. 6 deposited under Accession No. LMBP 3762.

120. Plasmid pLM4 of Sequence ID No. 7 deposited under Accession No. LMBP 3763.

121. Plasmid pEGF72 of Sequence ID No. 8 deposited under Accession No. LMBP 3764.

122. Plasmid pCB501 of Sequence ID No. 9 deposited under LMBP Accession No. LMBP 3765.

123. A worm strain comprising a chimeric C.elegans human unc-53 gene deposited under LMBP Accession No. LMBP-1663CB.

124. A vertebrate homologue according to any of claims 1 to 3 which is a mouse homologue.

125. A homologue according to claim 125 having the sequence illustrated in Figure 14.

1/270

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 16

AlwI

ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCA 9697
TGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTACGACAGAAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGT
P R S A N P V T S G C C Q V R . V V S Y R V G L K T I V T G . G A

ApaI

GCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC 9700
CGCCAGCCCGACTTGCCCCCAAGCAGTGTGTGGGTCGAACCTCGCTTGTCTGGATGTGGCTTGACTCTATGGATGTGCGACTCGATACTCTTTCGCGG
A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K R

ACGCTTCCCAGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATC 9703
TGCGAAGGGCTTCCCTCTTTCGCGCTGTCCATAGGCCATTGCGCGTCCCAGCCTTGTCTCTCGCGTGCCTCCCGAAGGTCCCCCTTTGCGGACCATAA
H A S R R E K G G Q V S G K R Q G R N R R A H E G A S R G K R L V S

TTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGC 9706
AAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGCGTCTGTCGCGC
L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R G

Ava III
Nsi I

CTTTTACGGTTCTTGCCCTTTTGTGCTGCTGCTTCTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 9709
GAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTAACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Tuesday, 18 November 1997 11:46

fig 31 pEGFPsma (1 > 6960) Site and Sequence

Enzymes: 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACSGTAAATGSCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTGCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTCGCGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATCGGGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACCTACGCCAAAACCGTCATGTAGTTACCCGACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H O V A V
TAGCGGTTTGACTCACGGGGATTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGTTTGGCACAAAATCAACGGGACTTTCCAAATGTCGTA
ATCGCCAAACTGAGTGCCCTAAAGTTAGAGGTGGGTAACGTCAGTTACCTCAAAACAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCA
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S
ACAACCTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACGCGTTTACCGCCATCCGCACATGCCACCTCCAGATATATTGCTCGACCAAACTACTGGCAGCTAGCGCATCGCGA
O L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTACGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACACGGGTAGGACCAGCTCGACCTGCCGCTGCATTGCGCGGTGTTCAAGTCG
eGFPc.e.unc53sma
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCGGCGGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCACTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCAACCTCGTGAC
ACAGGCGCTCCCGCTCCCGCTACGGTGGATGCCGTTGACTGGGACTTCAAGTAGACGTGGTGGCGTTTCGACGGGCACGGGACCGGTGGGAGCACTG
eGFPc.e.unc53sma
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
CACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTCTGAAGAAGTTACGGCGGTACGGGCTTCGGATGCAGGTCCTC
eGFPc.e.unc53sma
T L T Y G V O C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTGGTGAAACCGCATCGAGCTGAAGGCCA
GCGTGGTAGAAGAAGTTCCGTGCGCTTGATGTTCTGGGCGCGGC TCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGAATTCCCGTAGC
eGFPc.e.unc53sma
P T I F F K D D G N Y K T R A E V K F E G D I L V H R I E L K G I

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 2

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TACAAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCTCTCGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTCTCTTGGCGTAGTT 1100

eGFPc.e.unc53sma

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K O K N G I I

GGTGAACITCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGTTCGCCGACCCTACCAGCAGAACACCCCATCGGGCAGCGCCCGTGTCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCTGCCGTCGCACGTCGAGCGGCTGGTGTGTCGTC TTGTGGGGTAGCCGCTGCCGGGGCACGACGAC 1200

eGFPc.e.unc53sma

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGTGAGTGGACTCGTGGGTGAGGCGGGACTCGTTTCTGGGGTGTCTCTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGGCGGCCCT 1300

eGFPc.e.unc53sma

P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTTCAGGCCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCTTAACCCGGTTAGCCGTGGAAAG 1400

eGFPc.e.unc53sma

C.e.unc53 sma

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT 1500

eGFPc.e.unc53sma

C.e.unc53 sma

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCCTGCATTACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCTCGAAACGTGTC TCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTGTAGTGTAGCTTGGACCTACCGGAGCTTTCACAGAGCTGATGGACTTTTGTAGACCCAGAGCTGACGA 1600

eGFPc.e.unc53sma

C.e.unc53 sma

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L D C

CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTC TCCAGCTGCTCTTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTGTAGTGGTTTTGGCTATAGCTGTCGCCCTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCTGCTTTCGAAGCCGTTGACTT 1700

eGFPc.e.unc53sma

C.e.unc53 sma

S K L T K T D I D S S Y L G A V L Q L L F L L S T Y K O K L R Q L I

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 1

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTCTAAATTACCCTCGCCACGTGTCGCCACGTACAGCAACCGCTTCA
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCTGTTGGCGAAGT 1800

eGFPc.e.unc53sma

C.e.unc53 sma

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCACCTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCGGAAGTCTGAGGTGTCAGTTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT 1900

eGFPc.e.unc53sma

C.e.unc53 sma

A T N P N S N F P Q M S T S R L Q T P O S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCTCATCATCAACCACTTCATCAATAATACAAATTCATTCCGTCGTCGAGCCGTTTCGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT 2000

eGFPc.e.unc53sma

C.e.unc53 sma

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTTGAATCTAAACCGACCTACCTCCCAACTCCAAAAA
ACAACCGAGCTGCTATAGGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTGGCTGGATGGAGGGTTGAGGTTTTT 2100

eGFPc.e.unc53sma

C.e.unc53 sma

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I

CCTTCTAGACCACAAACCCAGCTAGTTCTGTTGCTACAACACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAACTT
GGAAGATCTGGTGTGTTGGGTCGATCAAGCACACGATGTTGATGTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACCTCGTGGGGTTTTSAAC 2200

eGFPc.e.unc53sma

C.e.unc53 sma

P S R P Q T Q L V R V A T T T K I G S S K L A A P K A V S T P K L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCCGATAACAGCGGTGGTGGTGGTGGGAATGCTGAAATTAAGTTATTAGTAGCAAAAACCCATC
GAAGACACTTCTGATAACCTCGTTTGTCTCGGGCTATTGTCGCCACCACCACCACCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG 2300

eGFPc.e.unc53sma

C.e.unc53 sma

A S V K T I G A K O E P D N S G G G G G G M L K L K L F S S K N P S

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 4

TTGCTCATCGAATAGCCCAACCTACGAGAAAGCGGCGGCGGTGCC TCAACAACAACTTTGTGCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AAGGAGTAGCTTATCGGGTGTGGATGCTCTTTCCGCCGCCGCCACGGAGTTGTTGTTGAAAACAGCTTTTAGCGACGGGGTCACTTTCACCGGACTTC 2400

eGFPc.e.unc53sma

C.e.unc53 sma

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L I

CCGCCGACCAGTAAGCTGGGAAGTGCCACGCTATGTGGAAGCTTTGTACGCCAAAGTTTCCTACCGTAAAACGGACGCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTGACCCCTTACGGTGCAGATACAGCTTCGAAACATGCGGTTTCAAAGGATGGCATTTCCTGCGGGGTTAGTATAGAGTTGTTTC 2500

eGFPc.e.unc53sma

C.e.unc53 sma

P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q

ACTCGAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTCGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTTCTCGTCACTTCTTCTCAGGCCATGCGACCTAAGTTGCTGTCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTTCGTA 2600

eGFPc.e.unc53sma

C.e.unc53 sma

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCCTCCATCGTGACAGCTATCAGACAGCGG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGTAGTAATGAGAATTGCGGAGGTAGCACGTGCGATAGTCTGTGCGG 2700

eGFPc.e.unc53sma

C.e.unc53 sma

F S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

AAGCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAAGATC
TATCGGCGTGTGGCCAAAGAGGTTTATAATAGTTGTTGCGACAACCTCTTTTGGTTGTGACCGTCACTTTCTCACTTTTCGTGTCGCTTTTCTAG 2800

eGFPc.e.unc53sma

C.e.unc53 sma

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTGACAAAGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGATTCTTCAACTGTTTACTGGGGCACTATAGACT 2900

eGFPc.e.unc53sma

C.e.unc53 sma

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 8980) Site and Sequence

Page 6

AAAACCAAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAG 300

eGFPc.e.unc53sma

C.e.unc53 sma

K P E P E K L Q S M S I D T T D V P P L P P L K S V V P L K M T S

ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCGGCTGGAAGACT
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTCTCTTTTAGTGTAGCGGACAGTTTCAAGAACTATATCTGTCAGCAGGCGCAGACTTCTGA 310

eGFPc.e.unc53sma

C.e.unc53 sma

I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D

CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTGAAGACCATAGTAAGCGACCTCTCTTCTACCTTTCTTATTCGTAGTCTTAS 320

eGFPc.e.unc53sma

C.e.unc53 sma

S I V A H A S A O V T P P T K T S G N H S L E R R M G K N K T S E S

CASCGGCTACACCTCTGACGCCGGTGTGGGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT
GTGCGCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTACTCCCTCTTCGACTTTCTTATGCTACTGTACTGAGCAGCTCGTGCTTGCCGATA 330

eGFPc.e.unc53sma

C.e.unc53 sma

S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A Q N G Y

CCTGACAACCTCGAAGACAGTTCTCTTGTGCTGCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGS
GGACTGTTGAAGCTTCGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCTTAAACAGGCCATCTGTATAC 340

eGFPc.e.unc53sma

C.e.unc53 sma

P D N F E D S S S L S S G I S D N N E L D D I S T D D L S G V D P

CAACAGTCGCTTCCAAACATAGCGACTATTCCCACTTGTTCGCCATCCACGCTTCTTCTCAAGGCCCGAGTCCCCAGTCGGTCTCCACATCACT
GTTGTCAGCGAGGTTTGTATCGCTGATAAGGGTGAACAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTGTAGTCA 350

eGFPc.e.unc53sma

C.e.unc53 sma

A T V A S K H S D Y S H F V R H P T S S S S K P R V P S R S S T S V

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 6

CGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAAACTTCTGCCAGTGCCGAACGAGCCAACGTGGCGCCGCTGCCACCTCAACCTTCGGACAA
GCTAAGAGCTAGAGCTCGTCTTGCTCTTACACATGTTTGAAGACAGGGTCACGGCTTGCTCGGTTGCACCGGGCGACGGTGGAGTTGSAAGCCTGTT

360

-----aGFPc.e.unc53sma-----

-----C.e.unc53 sma-----

D S R S R A E Q E N V Y K L L S Q C R T S Q R G A A A T S T F G Q

CATTTCGCTAAGATCCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTACTTGCTTTAAAAAACCTCCAC
GTAAGCGATTCTAGGGGCCCTAGGTGGCTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAACATCTCCAAAATGAACGAAATTTTTTGGAGGGTG

370

-----aGFPc.e.unc53sma----->

-----C.e.unc53 sma----->

H S L R S P G S T G S R . L I I I S H T T F V E V L L A L K N L P

ACCTCCCCCTGAACCTGAACATATAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAA
TGGAGGGGACTTGGACTTTGTATTTTACTTACGTTAACAACAACAAATGAACAAATAACGTGGAATATTACCAATGTTTATTTCTGTTATCGTAGTGT

380

H L P L N L K H K M N A I V V V N L F I A A Y N G Y K . S N S I T N

TTTCACAAATAAAGCATTTTTTCTACGCTTCTAGTTGTGGTTTGTCAAACTCATCAATGTATCTTAACGCGTAAATTTGAAGCGTTAATATTTGTT
AAAGTGTATTATTTCTGTAATAAAGTACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATTCGCGATTAAACATTCGCAATTATAAAACAA

390

F T N K A F F S L H S S C G L S K L I N V S . R V N C K R . Y F V

AAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAGAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGG
TTTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATTGGTTATCCGGCTTTAGCCGTTTAGGGAATATTTAGTTTCTTATCTGGCTCTATCCC

400

K I R V K F L L N Q L I F . P I G R N R O N P L . I K R I D R D R

TTTSGTGTGTTCCAGTTTGAACAAGASTCCACTATTAAGAAGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCACTAC
AACTCACAACAAGGTCAAACCTTGTCTCAGGTGATAATTTCTGCACCTGAGGTTGACGTTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATG

410

V E C C S S L E O E S T I K E R G L Q R Q R A K N R L S G R W P T T

GTGAACCATCACCTTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTTAAGGGAGCCCCGATTAGAGCTTGACGGGGA
CACTTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCGTGATTAGCCTTGGGATTTCCTCGGGGGCTAAATCTGAACATGCCCTTT

420

. T I T L I K F F G V E V P . S T K S E P . R E P P I . S L T G I .

GCCGGCGAACGTGGCGAGAAAGGAAGGAAGGAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACAGCTGCGCGTAACCAACACACCC
CGGGCGCTTGACCGCTCTTTCTTCCCTTCTTTCGCTTTCCTCGCCCGCATCCCGGACCGTTACATCGCCAGTGGCAGCGCATTGGTGGTGGG

430

A G E R G E K G R E E S E R S G R . G A G K C S G H A A R N H H T

GCCGCGCTTAATGCGCCGTACAGGGCGGTACAGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAAATACATTCAAATA
CGGCGGAATACGCGCGGATGTCCCGCGCAGTCCACCGTGAAAGCCCCCTTACACGCGCTTGGGGATAAACAAATAAAAGATTATGTAAAGTTTAT

440

R R A . C A A T G R V R W H F S G K C A R N P Y L F I F L N T F K Y

TGATCCGCTCATGAGACAATAACCTTGATAAATGCTTCAATAATATTGAAAAAGGAGAGTCTGAGGCGGAAAGAACAGCTGTGGAATGTGTGTCAG
ACATAGGCGAGTACTCTGTATTGGGACTATTACGAAGTTATTATAACTTTTCTCTCAGGACTCCGCTTCTTGGTCGACACCTTACACACAGTC

450

V S A H E T I T L I N A S I I L K K E E S . G G K N Q L V N V C Q

Tuesday, 18 November 1997 11:48
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 7

TTAGGGTGTGGAAAGTCCCAGGCTCCCAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAGGCTCCC
AATCCCACACCTTTCAGGGGTCCGAGGGGTCTCCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGG
L G C G K S P G S P A G R S M Q S M H L N . S A T R C G K S P G S 280

CAGCAGGCAGAAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACCCGCCATCCCGCCCTAACCCGCCAGTTCGGC
GTCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGC
P A G R S M Q S M H L N . S A T I V P P L T P P I P P L T P P S S A 270

CCATTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAAGAGGCCAGGGCCCTCGGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTT
GGTAAGAGGGCGGGTACCGACTGATTAATAAATAAATACGCTCCGGCTCCGGCGAGCCGAGACTCGATAAGGTCTTCATCACTCTCCGAAAAA
H S P P H G . L I F F I Y A E A E A S A S E L F Q K . . G G F F . 280

GGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTCTCCGGCCGCTTGG
CCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTAGCAAGCGTACTAATTTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCGAACC
G G L G F C K D R S R D R M R I V S H D . T R V I A R R F S G R L 490

GTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCCGCTGTTCCGGCTGTGACGCGAGGGGCGCCGGTCTTTTTTG
CACCTCTCCGATAAGCCGATACTGACCCGTGTGTCTGTGTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCGCGTCCCGGGGCAAGAAAAAC
G G E A I R L . L G T T D N R L L . C R R V P A V S A G A P G S F C 500

TCAAGACCGACCTGTCCGGTGCCTGAATGAAGTCAAGACGAGGCAGCGGGCTATCGTGGCTGGCCACGACGGCGTTCCTTGCGCAGCTGTGCTCGA
AGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCAGCGGTGCTGCCCGCAAGGAACGCGTCGACAGAGCT
O D R P V R C P E . T A R R G S A A I V A G H D G R S L R S C A R 510

CGTTGTCACTGAAGCGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCTGTCTCTCACCTTGCTCTCCGAGAAAGTATCCATC
GCAACAGTGACTTCGCCCTTCCGTGACCGACGATAACCCGCTTACGCGCCCGTCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTCATAGGTAG
R C H . S G K G L A A I G R S A G A G S P V I S P C S C R E S I H 520

ATGGCTGATGCAATCGCGGGCTGCATACGCTTGATCCGGCTACCTGCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGG
TACCGACTACGTTACGCGCCGACGTATGCGAATAGGCGGATGGAGGGTAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGATGAGCCTACC
H G . C N A A A A Y A . S G Y L P I R P P S E T S H R A S T Y S D G 530

AAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGCGA
TTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTGCTAGTCCCGAGCGGGTGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACGGGCTGCGGT
S R S C R S G . S G R R A S G A R A S R T V R Q A Q G E H A R R R 540

GGATCTCGTCGTGACCCATGGCGATGCCGTGCTTGCCGAATATCATGGTGGAAATGGCCGCTTTTCTGGATTATCGACTGTGGCCGGCTGGGTGTGGG
CCTAGAGCAGCACTGGGTACCGTACTGACGACGAACGGCTTATAGTACCACCTTTTACCGGCGAAAGACCTAAGTAGCTGACACCGGCCACACACCGC
G S R R D P V R C L L A E Y H G G K V P L F V I H R L V P A G C G 550

GACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTC
CTGGCGATAGTCTGTATCGCAACCGATGGGCACATAACGACTTCTGCAACCGCGCTTACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAG
G P L S G H S V G Y P . Y C . R A V R R M G . P L P R A L R Y R R S 560

CCGATTTCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGGCA
GGCTAAGCGTCCGCTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTACTGGCTGGTTCGCTGCGGTTGGACGCT
R F A A H R L L S P S . R V L L S G T L G F E M T D Q A T P N L P 570

Tuesday, 18 November 1997 11:48
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 4

TCACGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
AGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGGCGACCTACTAGGAGGTCGCGCCCCTAGT
S R D F D S T A A F Y E R L G F G I V F R D A G W M I L Q R G D L 580

TGCTGGAGTTCTTCGCCACCTAGGGGAGGCTAACTGAAACACGGAAGGAGACAATACGGAAGGAACCGCGCTATGACGGCAATAAAAGACAGAA
ACGACCTCAAGAAGCGGGTGGGATCCCCCTCGATTGACTTTTGCCCTTCTCTGTTATGGCCTTCTTGGGCGCGATACTGCCGTTATTTTCTGTCTT
M L E F F A H P R G R L T E T R K E T I P E G T R A M T A I K R Q H 590

TAAACGCACGGTGTGGGTCGTTTTCATAAACCGGGGTTGGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCAATTGGGGCCAATAC
ATTTTGGTGGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCAGCGTGAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATG
K T H G V G S F V H K R G V R S Q G V H S V D T P P R P H V G Q Y 600

GCCGCGTTTCTTCTTTTCCCAACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCAGCCAACGTCGGGCGGCAGGCCCTGCCATAGCCTCAG
CGGGCGCAAGAAGGAAAAGGGTGGGGTGGGGGTTCAAGCCACTTCGGGTCCCGAGCGTCGGTTGCAGCCCCGCGTCCGGGACGGTATCGGAGTC
A R V S S F S P P H P P S S G E G P G L A A N V G A A G P A I A S 610

GTTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTA
CAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTCTAGATCCACTTCTAGGAAAACTATTAGAGTACTGGTTTTAGGGAAT
G Y S Y I L . I D L K L H F . F K R I . V K I L F D N L M T K I P 620

ACGTGAGTTTTCTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAACA
TGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGCATCTTTCTAGTTCTTAGAAGAACTCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTGT
R E F S F H . A S D P V E K I K G S S . D P F F L R V I C C L Q T 630

AAAAAACCCGCTACCAGCGGTGGTTTGTGTCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGCAGATACCAATA
TTTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAGGCTTCCATTGACCGAAGTCGTC TCGGCTCTATGGTTAT
K K P P L P A V V C L P D Q E L P T L F P K V T G F S R A Q I P H 640

CTGTCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGC
GACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTCTTGAGACATCTGGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACG
T V L L V . P . L G H H F K N S V A P P T Y L A L L I L L P V A A A 650

CAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGTGGGCTGAACGGGGGGTTCGTGCACACAGCC
GTCACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTC TGCATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCAAGCACGTTGTGCGG
S G D K S C L T G L D S R R . L P D K A Q R S G . T G G S C T Q P 660

AGCTTGGAGCGAACGACCTACACCGAAGTGAATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGGCGGACAGGTATCGGG
TCGAACCTCGCTTGTGGAATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCGGGTGCGAAGGGCTTCCCTCTTTCGGCTGTCCATAGCC
S L E R T T Y T E L R Y L Q R E L . E S A T L P E G R K A D R Y P 670

TAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAACGCCTGGTATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGA
ATTGCGCGTCCAGCCTTGTCTCTCGGTGCTCCCTCGAAGGTCCCCCTTTCGGGACCATAGAAATATCAGGACAGCCCAAGCGGTGGAGACTGAAC
V S G R V G T G E R T R E L P G G H A V Y L Y S P V G F R H L . L E 680

GCSTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCCTGGCTTTTGTGCGCTTTTGTG
CGCAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTGGGTCGTTGCGCGGAAAAATGCCAAGGACCGGAAAAACGCGGAAAAACGA
R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F A 690

Tuesday, 18 November 1997 11:46
fig 31 pEGFPsma (1 > 6960) Site and Sequence

Page 9

• CACATGTTCTTTCC TGC GTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
GTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 6960
H M F F P A L S P D S V D N R I T A M H

Tuesday, 18 November 1997 11:46

fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Enzymes : 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

3p.

TAGTTATTAATAGTAATCAATACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATGAATGCCATTACCGGGCGGACCGACTGGG
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTCAGCTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGGCGGTAACGTCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGTCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACGTCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGG
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCCGCACT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H O V A W
TAGCGGTTTGACTCACGGGATTTCGAAGTCTCCACCCCATTCAGCTCAATGGGAGTTGTGTTTGGCACCAAAATCAACGGGACTTTCAAAATGTCTGA
ATCGCCAAATGAGTGCCCTAAAGGTTGAGAGGTGGGGTAACGTCAGTTACCTCAACAAAACCGTGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .
ACAATCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGCGGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTGCTCTGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT
Q L R P I D A N G R . A C T V G G L Y K O S V F S E P S D P L A L
CCSGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTCCCATCTGGTCGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCG
GGCCAGCGTGGTACCACTCGTTCCCGCTCTCTGACAAGTGGCCCCACCAAGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGGTGTCAAGTCGG
eGFPC.e.unc53ec1
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCACTGCAACACCGGCAAGCTGCCCGTGCCCTGGCCACCTCTGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCGGTTTCGACGGGACGGGACGGGTGGGAGCACTG
eGFPC.e.unc53ec1
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
CACCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCGACACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCGGAAGGCTACGTCCAGGAG
GTGGGAC TGGATGCCGACGTACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCTGTCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCTCT
eGFPC.e.unc53ec1
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A P P E G Y V Q E
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GGTGGTAGAAGAAGTTCCTGCTGCCGTGTATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCACCTGGCGTAGCTCGACTTCCCGTAGG
eGFPC.e.unc53ec1
P T I F F K D D G N Y K T R A E V K F E G D T L V H R I E L K G I

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 2

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAG
TGAAGTTCCCTCCGCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCAGGCTGTTCTGCTCTCTTCCCGTAGT

eGFPC.e.unc53ed1

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I

GGTGAACATCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCCTGCGTCGACGTCGAGCGGCTGGTGATGGTCTGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC

eGFPC.e.unc53ed1

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCCTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTGGTGATGGACTCGTGGGTACGGCGGGACTCGTTTCTGGGGTGTCTCTCGCGCTAGTGACCAGGACGACCTCAAGCACTGGCGGGGGCCCT

eGFPC.e.unc53ed1

P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTT
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGATGCAAGTTACATCTTAACATATGGTTAGATGTGCTTAACCCGTTAGCCGTGGAAAG

eGFPC.e.unc53ed1

C.e.unc53 ed

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D W A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACATAGCAAGGCTAGTTGCT

eGFPC.e.unc53ed1

C.e.unc53 ed

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCTGCAATTCACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAAACGTTTTAGTGTAAGCTTGGACCTACCGGAGCTTGCACAGAGCTGATGGACTTTTGTAGACCCAGAGCTGACGA

eGFPC.e.unc53ed1

C.e.unc53 ed

F S P A F T K R L A K I T S H L D G L E T C L D Y L K N L G L D C

CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGTGCTCTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGTGTCGCCTTTGAACCCACGTCAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCTGCTTTCGAAGCCGTTGACT

eGFPC.e.unc53ed1

C.e.unc53 ed

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 3

AAAAGATCAGAAGAAATTGGAGCAAC TACCCACATCCATTATGCCACCCGCGGTTCTAAATTACCTCGCCACGTGTCGCCACGTACAGCAACCGCTTCA
TTTTCTAGTCTTCTTTAACCTCGTTGATGGGTGATAGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCTGTTGGCGAAG

eGFPC.e.unc53ec1

C.e.unc53 ed

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCTCACTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAAATATCGAAAATTGATTCATCAAGATTGGTATCA
CGTTGATTGGGTTTAAGGTTGAAAGGTGTTACAGTTGTAGGTCCGAAGTCTGAGGTGTCAGTTCTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT

eGFPC.e.unc53ec1

C.e.unc53 ed

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCGGTCGAGCCGTTGAGTGGCAATAATAA
TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTAT

eGFPC.e.unc53ec1

C.e.unc53 ed

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N H

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAA
ACAACCGAGCTGCTATAGGTGTAGACGCTTCTCGAATCTTAGTAGTAGTTGCATGTCGAGATAAAGCTTAGATTGGCTGGATGGAGGGTTGAGGTTTTT

eGFPC.e.unc53ec1

C.e.unc53 ed

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I

CCTTCAGACCACAAACCCAGCTAGTTCTGTGTTGCTACAACCTACAAAAATCGGAAGCTCAAAGCTAGCCGCTCCGAAAGCCGTGAGCACCCCAAACTTG
GGAAGATCTGGTGTGTTGGGTGATCAAGCACACGATGTTGATGTTTTAGCCTTCGAGTTTCGATCGGCGAGGCTTTCGGCACTCGTGGGGTTTTGAAC

eGFPC.e.unc53ec1

C.e.unc53 ed

P S R P Q T Q L V R V A T T T K I G S S K L A A P K A V S T P K L

CTTCTGTGAAGACTATTGGAGCAAAACAAGAGCCGATAACAGCGGTGGTGGTGGTGGGAATGCTGAAATTAAGTTATTTCAGTAGCAAAAACCATC
GAAGACACTTCTGATAACCTCGTTTGTTCGCGGCTATGTCGCCACCACCACCACCCTTACGACTTTAATTTCAATAAGTCATCGTTTTTGGGTAG

eGFPC.e.unc53ec1

C.e.unc53 ed

A S V K T I G A K Q E P D N S G G G G G G M L K L K L F S S K N P S

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 4

TTCTCATCGAATAGCCCAACCTACGAGAAAGGCGCGGGTGCCTCAACAACAACTTTGTCGAAAATCGCTGCCCCAGTGAAAAGTGGCCTGAAG
AAGGAGTAGCTTATCGGGTGTGGATGCTCTTCCGCGCGCCACGGAGTTGTTGTTGAAACAGCTTTAGCGACGGGGTCACTTTACCGGACTTC

eGFPC.e.unc53ed

C.e.unc53 ed

S S S N S P Q P T R K A A A V P Q Q Q T L S K I A A P V K S G L A

CGCCGACCAAGTGAAGTGGGAAGTGCACGTCTATGTCGAAGCTTTGTACGCCAAAAGTTCTACCGTAAAACGGACGCCCAATCATATCTCAACAAG
GGCGGCTGGTCATTGACCTTCACGGTGCAGATACAGCTTCGAAACATGCGGTTTTCAAAGGATGGCATTTCCTGCGGGGTAGTATAGAGTTGTC

eGFPC.e.unc53ed

C.e.unc53 ed

P P T S K L G S A T S M S K L C T P K V S Y R K T D A P I I S Q Q

ACTCGAAACGATGCTCAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCACGTGCCAACGTCATCATCGACGGAAGGTTCCCTAAGCAT
TGAGCTTTGCTACGAGTTCTCGTCACTTCTCTCAGGCCATGCGACCTAAGTTGCTGTCAGCGGTTGCAGTAGTAGCTGCCTTCCAAGGGATTCTGA

eGFPC.e.unc53ed

C.e.unc53 ed

D S K R C S K S S E E E S G Y A G F N S T S P T S S S T E G S L S M

GCATTCCACATCTTCCAAGAGTTCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCCGCCATCGTGACAGCTATCAGACAGCGG
CGTAAGGTGTAGAAGGTTCTCAAGTTGCAGTCTGCTTTTCAGAGGCAGTAGTCTGCTAGAATGAGAATTGCGGAGGTAGCACTGTGATAGTCTGTGCGG

eGFPC.e.unc53ed

C.e.unc53 ed

H S T S S K S S T S D E K S P S S D D L T L N A S I V T A I R Q P

ATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACTGGCAGTGAAAGGAGTGAAAAGCACAGCGAAAAAGATC
TATCGGCGTTGTGGCCAAAGAGGTTTATAATAGTTGTTGCGACAACCTCTTTTGTTGTTGACCGTCACTTTCCTCACTTTCGTGTCGCTTTTTTCTAG

eGFPC.e.unc53ed

C.e.unc53 ed

I A A T P V S P N I I N K P V E E K P T L A V K G V K S T A K K D

CACCTCCAGCTGTTCCGCCACGTGACACCCAGCCAACAATCGGAGTTGTTAGTCCAATTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGA
GTGGAGGTCGACAAGGCGGTGCACTGTGGGTCGGTTGTTAGCCTCAACAATCAGGTTAATACCGTGATTCTTCAACTGTTTACTGGGGCATATAGACT

eGFPC.e.unc53ed

C.e.unc53 ed

P P P A V P P R D T Q P T I G V V S P I M A H K K L T N D P V I S E

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 6

AAAACCAAGAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTTCCACCTCTAAAATCAGTTGTTCCACTTAAAATGACTTCA
TTTGGTCTTGGACTTTTCGAGGTTAGTTACTCGTAGCTGTGCTGCCTGCAAGGTGGCGAAGGTGGAGATTTTAGTCAACAAGGTGAATTTTACTGAAGT
eGFPC.e.unc53ec1
C.e.unc53 ed
K P E P E K L Q S M S I D T T D V P P L P P L K S V V P L K M T S
ATCCGACAACCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAGTCTTTGGATATGAGCAGTCGTCCGCTCTGAAGACT
TAGGC TGTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTAGTGTAGCGGACAGTTCAGCAAACCTATACCTCGTCAGCAGGCGCAGACTTCTGA
eGFPC.e.unc53ec1
C.e.unc53 ed
I R O P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D
CCATTGTGGCTCATCGCTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAACT
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTGAAGACCATTAGTAAGCGACCCTCTCTCTACCTTTCTTATTCTGTAGCTTAG
eGFPC.e.unc53ec1
C.e.unc53 ed
S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S
CAGCGGCTACACCTCTGACGCCGTTGTCGATGTGCGCCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAAGGCTAT
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTCTTATGCTACTGTACTGAGCAGCTCGTGTCTTGCCGATA
eGFPC.e.unc53ec1
C.e.unc53 ed
S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A Q N G Y
CCTGACAACCTCGAAGACAGTTCCTCCTTGTCGCTGGAATATCCGATAACAACGAGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACT
GGACTGTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCCCTTAGGTGGCTTAGATCTATTGACTAGTATTAGTCGGTATGG
eGFPC.e.unc53ec1
C.e.unc53 ed
P D N F E D S S S L S S G I S D N N E S I H R I . I T D H N Q P Y
ACATTGTAGAGGTTTTACTTGTCTTAAAAAACCTCCACACCTCCCTTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTATTATTS
TGTAAACATCTCCAAAATGAACGAAATTTTTGGAGGGTGTGGAGGGGACTTGGACTTTGTATTTTACTTACGTTAAACAACAATGAACAAATAAC
H I C R G F T C F K K P P T P P P E P E T . N E C N C C C . L V Y C
CAGCTTATAATGGTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACCTGCATTCAGTTGTGGTTTGTCCAAACTCATCA
GTGGAATATTACCAATGTTTATTTCGTTATCGTAGTGTITAAAGTGTITATTTGTAAGGAGTACGTAAGATCAACACCAACAGGTTTGAGTAGTT
S L . V L Q I K O . H H K F H K . S I F F T A F . L V F V O T H Q

Tuesday, 18 November 1997 11:46
fig 32 pEGFPed (1 > 6700) Site and Sequence

Page 6

TGTATCTTAACGCGTAAATGTGAAGCGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTAAACCAATAGGCCGAAATCGG
ACATAGAATTGCGCATTTAAACATTGCAATTATAAAACAATTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTAGCC 370
C I L T R K L . A L I F C . N S R . I F V K S A H F L T N R P K S
CAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCAGTTTGAACAAGAGTCCACTATTAAGAACGTGGACTCCAACGTC
GTTTATAGGAATATTTAGTTTCTTATCTGGCTCTATCCCACTCACAACAAGGTCAAACCTGTTCAGGTGATAATTTCTTGACCTGAGGTGCA3 380
A K S L I N O K N R P R . G . V L F Q F G T R V H Y . R T W T P T S
AAAGGCGAAAAACCGTCTATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCGGTAAAGCACTAAATCGG4
TTTCCGCTTTTGGCAGATAGTCCGCTACCGGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCGTGATTAGCCT 390
K G E K P S I R A M A H Y V N H P N O V F W G R G A V K H . I G
ACCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAGCCGGCGAAGTGGCGAGAAAGGAAGGAAGGAAGGAGCGGGCGCTAGGGCGCT
TGGGATTTCCCTCGGGGCTAAATCTCGAAGTGCCTTTTCGGCGCTTGACCGCTTTTCTTCCCTTCTTCGCTTTCCTCGCCGCGATCCCGCG4
T L K G A P D L E L D G E S R R T V R E R K G R K R K E R A L G R
GGCAAGTGTAGCGGTACGCTGCGCTAACCACACACCCGCGCTTAATGCGCGCTACAGGGCGCTCAGGTGGCACTTTTCGGGGAATGTGCGE
CCGTTACATCGCCAGTGGCAGCGCATTTGGTGGTGTGGGCGGCGCAATTACGCGCGATGTCCCGCGCAGTCCACCGTGAAGAGCCCTTTACACGCE 410
V O V . R S R C A . P P H P P R L M R R Y R A R O V A L F G E M C A
GGAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
CCTTGGGGATAAAACAAATAAAAGATTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATACTTTTCTTCT 420
E P L F V Y F S K Y I O I C I R S . D N N P D K C F N N I E K G R
GTCCTGAGGCGGAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGATGCAAGCATGCATCTC
CAGGACTCCGCTTTCTTGGTGCACACCTTACACACAGTCAATCCACACCTTTACAGGGTCCGAGGGGTGCTCCGCTTTCATACGTTTCTGACGTAGAG 430
V L R R K E P A V E C V S V R V V K V P R L P S R O K Y A K H A S
AATTAGTCAGCAACCAGGTGTGGAAGTCCCAGGCTCCCAGCAGGCAGAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
TTAATCAGTCGTTGGTCCACACCTTTACAGGGTCCGAGGGTCTGCTGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGG 440
Q L V S N Q V V K V P R L P S R O K Y A K H A S Q L V S N H S P A P
TAACTCCGCCATCCGCCCTAACTCCGCCAGTTCCGCCATTCTCCGCCCATGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTC
ATTGAGGCGGGTAGGGCGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAATAAATAAATACGCTCCTCGGCTCCGGCGGAG 450
N S A H P A P N S A Q F R P F S A P V L T N F F Y L C R G R G R L
GGCCTCTGAGCTATTCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTAGGCTTTTGAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGAT
CCGGAGACTCGATAAGGTCTTCACTCTCCGAAAAACCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGCTACTCTTAGCAAGCGTACTA 460
G L . A I P E V V R R L F W R P R L L Q R S I K R Q D E D R F A .
TGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCAACAACAGACAATCGGCTGCTCTGATGCCGC
ACTTGTCTTACCTAACGTGCGTCAAGAGGCGGCGCAACCCACCTCTCCGATAAGCCGATACGACCGGTGTGCTGTTAGCCGACGAGACTACGGCGG 470
L N K M D C T O V L R P L G V R G Y S A M T G H N R O S A A L M P P
GTGTTCGGCTGTACAGCGAGGGGCGCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGGCTGTAATGAACGTGAAGACGAGGACGCGGGCTATCGT
CACAAAGGCGACAGTCCGCTCCCGCGGGCAAGAAAAACAGTTCTGGCTGGACAGGCCAGGGACTTACTTGACGTTCTGCTCCGTCGCGCGGATAGGA 480
C S G C O R R G A R F F L S R P T C P V P . M N C K T R O R G Y F

Tuesday, 18 November 1997 11:46
fig 32 pEGFPec1 (1 > 6700) Site and Sequence

Page 7

GGCTGGCCACGACGGGCGTTCTTGCACAGCTGTGCTCGACGTTGCTACTGAAGCGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCT
CCGACCGGTGCTGCCCGCAAGGAACGCTGACACGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGATAACCCGCTTCACGGCCCCGCTCTAGA
G V P R R A F L A Q L C S T L S L K R E G T G C Y V A K C R G R I
CCTGTCTCTACCTTGCTCTGCGGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAE
GGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCGCGACGTATGCGAACTAGGCCGATGGACGGGTAAAGCTGGTG
S C H L T L L L P R K Y P S V L M Q C G G C I R L I R L P A H S T I
CAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGCGGTCTTGTGCTAGCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCG
GTTGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACAGCTAGTCTCTACTAGACCTGCTTCTCGTAGTCCCGAGCGCGGTGCGG
K R N I A S S E H V L G W K P V L S I R M I V T K S I R G S R Q P
AACTGTTTCGCGAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGGCGAATATCATGGTGGAAATAGGCCG
TTGACAAGCGGTCCGAGTTCGCTCGTACGGGCTGCCGCTCTAGAGCAGCACTGGGTACCCTACGGACGAACGGCTTATAGTACCACCTTTTACCGGC
N C S P G S R R A C P T A R I S S . P M A M P A C R I S V W K M A
CTTTTCTGGATTCATCGACTGTGGCGGCTGGGTGTGGCGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAA
GAAAAGACCTAAGTAGCTGACACCGGCGACCCACACCGCCTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCGCTT
A F L D S S T V A G V V V R T A I R T . R V L P V I L L K S L A A N
TGGGCTGACCGCTTCTCGTGCTTACGGTATCGCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCAGAAATGCCATAGCGGGGAGGGCTAAGCGTCCGCTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA
G L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S
GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGATTTGATTCCACCGCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTC
CCCCAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTGTCTTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAGCCTTAGCAAAAG
G V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F
CGGGACCGCGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCCCTGCGGCGACCTACTAGGAGGTGCGGCCCC TAGAGTACGACCTAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATG
P G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N I
CGGAAGGAACCCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTGGGTCGTTTGTTCATAAACCGGGGTTGGTCCCAGGGCTGGCA
GCC TTCTTGGGCGGATAGTCCGCTTATTTTCTGTCTTATTTGCGTGCCACAACCCAGCAAAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGT
G R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A
CTCTGTGATACCCACCGAGACCCATTGGGGCCAATACGCCCGCTTTCTTCTTTTCCCCACCCCAAGTTGCGGTGAAGGCCAGGGCTC
GAGACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTATGCGGGCGCAAAGAAGGAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAG
L C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A
GCAGCCAACGTCGGGGCGGAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTTAAAACCTTATTTTAAATTTAAAGGATCTAGS
CGTCGGTTCAGCCCCCGCTCCGGGACGGTATCGGAGTCCAATGAGTATATGAAATCTAACTAAATTTTGAAGTAAAAATTAATTTTCTAGATCC
R S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . N D L G
TGAAGATCTTTTGTAAATCTCATGACCAAAATCCCTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTTTTS
ACTTCTAGGAAAACTATTAGAGTACTGGTTTATGGGAATTGCACTCAAAGCAAGGTGACTCGCAGTC TGGGGCATCTTTCTAGTTTCTAGAAAGAC
E O P F . . S H D O N P L T . V F V P L S V R P R R K D O R I F L

Tuesday, 18 November 1997 11:46
fig 32 pEGFPect (1 > 6700) Site and Sequence

Page 9

AGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAACCACCGCTACCAGCGGTGGTTTGTGTCGGGATCAAGAGCTACCAACTCTTTTT
TCTAGGAAAAAAGACGCGCATTAGACGACGAACGTTTGTGTTTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 3100
R S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F
CCSAAGGTAAC TGCTT CAGCAGAGCGCAGATACCAATAC TGCTTCTAGTGTAGCCGTAGTTAGGCCACCACCTCAAGAACTCTGTAGCACC GCCTA
GGCTTCCATTGACCGAAGTCGCTCGCGCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGAT 3200
S E G N V L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y
CATACCTCGCTCTGCTAATCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCCTATTTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCG 3300
I P R S A N P V T S G C C Q V R . V V S Y R V G L K T I V T G . G
GCAGCGGTGCGGGTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGC
CGTCGCCAGCCGACTTGCCCCCAAGCAGTGTGTGCGGTCGAACCTCGCTTGTCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTG 3400
A A V G L N G G F V H T A Q L G A N D L H R T E I P T A . A M R K
GCCACGCTTCCC GAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC TGGT
CGGTGCGAAGGGCTTCCCTCTTTCCGCTGTCCATAGGCCATTCCGCGTCCAGCCTTGTCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCA 3500
R H A S R R E K G G Q V S G K R O G R N R R A H E G A S R G K R L V
ATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
TAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGAGCTAAAAACACTACGAGCAGTCCCCCGCTCGGATACCTTTTTCGCGTCTGTGCG 3600
S L . S C R V S P P L T . A S I F V M L V R G A E P M E K R Q Q R
GGCCTTTTACGGTTCCTGGCCTTTTGTGCGCCTTTTGTCTACATGTTCTTTCTGCGTTATCCCTTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CCGGAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGT 3700
G L F T V P G L L L A F C S H V L S C V I P . F C G . P Y Y R H A

Tuesday, 18 November 1997 11:47

fig 33 pEGFPxba (1 > 5447) Site and Sequence

Enzymes : 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

TAGTTATTAATAGTAATCAATACGGGGTCATTAGTTACATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCT
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAACGCAAGTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACGCAAGTATCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCTTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATCGGGGGGATAACGCAAGTATTGCAATTTACGGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCCTACGCCAAAACCGTCATGTAGTTACCGGCACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H O V A V
TAGCGGTTTGAATCAGGGGATTTCAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAATGTCGTA
ATCGCCAAATGAGTGCCCTAAAGGTTTCAAGGTTGAGGTTAAGTGCAGTTACCTCAACAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .
ACAACCTCCGCCCCATTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGGCGGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCTCCAGATATATTGCTCTGACCAAAATCACTTGGCAGTCTAGGCATCGCGAT
D L R P I D A N G R . A C T V G G L Y K Q S W F S E P S D P L A L
CCGGTCCGACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCTGGTTCGAGCTGGACGGCGACGTAACGGCCACAGTTTCAGCG
GGCCAGCGGTGGTACCACTCGTTCCCGCTCCTCGACAAGTGGCCCCACACGGGTAGGACCAAGCTCGACCTGCCGCTGCATTGCGCGGTGTTCAAGTGGC
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCCGGCGAGGGCGAGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTTATCTGCACCAACCGGCAAGCTGCGCGTGGCCGGCCCTCGTGAC
ACAGGCGCGTCCCGCTCCCGCTACGGTGGATGCCGTTGCACTGGGACTTCAAGTAGACGTGGTGGCCGTTGACGGGCGACGGGACCGGGTGGGAGCACTG
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P V P T L V T
CACCTGACCTACGCGGTGCAAGTCTTACCGCTACCCCGACCATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGAG
GTGGGACTGGATGCCGACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCTCT
T L T Y G V O C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
CGACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCGAGGTGAAGTTGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCSTGGTAGAAGAGTTCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 1

ACTTCAAGGAGGACGGCAACATCTGGGGCACAAAGCTGGAGTACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCTCTGCGGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTCTTGGCCGTAGTT

110

eGFPC.e.unc53xba

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I L

GGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTTGTAGCTCTGCGCTCGCACGTCGAGCGGCTGGTGATGGTCTGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC

120

eGFPC.e.unc53xba

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGCTGGAGTTCGTGACCGCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGATACAGGACGACCTCAAGCACTGGCGGGCGCCCT

130

eGFPC.e.unc53xba

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCTTAACCCGGT TAGCCGTGGAAAG

140

eGFPC.e.unc53xba

C.e.unc53 xba

I T L G M D E L Y K S G L R S T S N V E L I P I Y T D V A N R H L S

GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTCTCAGCTTATTAAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAGCTAATCCCTATAAAGGTTACTAAAGCGCTGATAGCTGACCAAGAGTCGAATAATTACAC TAGCAAGGCTAGTTGCTT

150

eGFPC.e.unc53xba

C.e.unc53 xba

K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

TTCTCGCCTGCATTACGAAACGTTTGGCAAAATCACATCGAACCTGGATGGCTCGAAACGTGTCTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTTGCAAACCGTTTTTAGTGAGCTTGGACCTACCGGAGCTTTCACAGAGCTGATGGACTTTTAGACCCAGAGCTGACGA

160

eGFPC.e.unc53xba

C.e.unc53 xba

F S P A F T K R L A K I T S N L D G L E T C L O Y L K N L G L D C

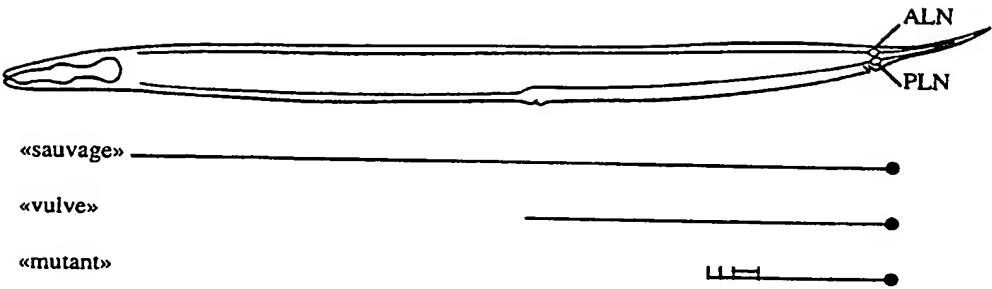
CGAAACTCACAAAACCGATATCGACAGCGGAAACTTGGGTGCAGTTCTCCAGCTGCTCTCTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTCGCTTTGAACCCACGTCAAGAGGTCGACGAGAAGGACGAGAGGTGGATGTTCTGCTTTCGAAGCCGTGACTT

170

eGFPC.e.unc53xba

C.e.unc53 xba

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R O L I



souches	phénotypes	sauvage	vulve	mutant	nombre
wt	ALN	100	0	0	70
	PLN	100	0	0	70
<i>unc-53(n152)</i>	ALN	0	26	74	50
	PLN	nd	nd	nd	
<i>unc-53(n152)</i> <i>pAΔunc-53</i>	ALN	95	0	5	65
	PLN	77	0	23	44
<i>unc-53(n152)</i> <i>pAΔunc-53-H1</i>	ALN	25	60	15	56
	PLN	52	16	32	25

Figure 51b

Tuesday, 18 November 1997 13:58

fig 52 pLM5 (1 > 5425) Site and Sequence

Page 1

Enzymes : All 148 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

GACGGATCGGGAGATCTCCCGATCCCCATGGTGCAGTCTCAGTACAATCTGCTCTGATGCCGATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCC TAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTTCATAGACGAGGGACGAACACACA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGC CGGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAAGTGGCTGTTAAGCTACTTCTTAGACGAATCCCAATCCGCAAAACGC
G G R . V V R E O N L S Y N K A R L D R O L H E E S A . G . A F C

CTGCTTCGCGATGACGGGCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGTCTATATGCGCAACTGTAACATAAATGATCAATAATTATCATTAGTTAATGCCCGAGTAATCAAGTATCGGGTATAT
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCGCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTACCGGGCGGACCGACTGGCGGGTGTCTGGGGGCGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAACGTGAGTTACCCACCTGATAAATGCCATTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGACAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTGATGTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAGGGGATTTCAGGCTCCACCCCATGACGTCAA
AGCGATAATGGTACCACTACGCCAAAACCGTCATGTAGTTACCCGCACCTATCGCCAAACTGAGTGCCCTTAAAGGTTACAGAGTGGGGTAACGTCAGTT
R Y Y H G D A V L A V H Q W A V I A V . L T G I S K S P P H . R Q

TGGGAGTTTGTGTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCCGCCCATTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAG
ACCTCAAAACAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTACAGCATGTTGAGGCGGGTAACGCGTTTACCCGCCATCCGCACATGCCACCCCTC
V E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGCTCGAGAGACCGATTGATCTCTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGGATCG
G L Y K O S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A

GTTTAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGATGGTATGGCTAGCATGACTGGTGGACAGCAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTCGAATGGTACCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACGACCACTGTCGTTTACCCAGCCCTAGACATGCTG
F K L K L T M G G S H H H H H H G M A S M T G G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 2

GATGACGATAAGGTACCTAGGATCCATGCAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGA
CTACTGCTATTCCATGGATCCTAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTTCTTTACT
pLM5 insert = U3
D D D K V P R I H A N E E E E P E K K E V S E L R S E L V E K E N
AGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGA4
TCGAATGCTGTAGGCGAACCCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACT
pLM5 insert = U3
ORF U3
K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L E V D L L A
AGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCACGCCGCTCC
TCGTCTCTTACTGGCTGACTTCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGGCAGG
pLM5 insert = U3
ORF U3
A E N D R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S
CTAGGCCCTGGCACTCACCATTCCTTCGGCCCCAGTCTTCGAGACACAGACCTGTCAACCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGA
GATCCGGACCGTGAGTGGGTAAGGAAGCCGGGGTCAGAAGCTCTGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCTCTCTTCA
pLM5 insert = U3
ORF U3
L G L A L T H S F G P S L A D T D L S P M D G I S T C G P K E E V
CCCTCCGGGTGGTGGTGAAGTACCCCCGAGCACATCATCAAGGGGACTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGCAAGGTGAGTGGAAAAGT
GGGAGGCCCAACCACTCTCTACGGGGCGTCTGTGAGTAGTTTCCCCGAACTTCGTCTCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCA
pLM5 insert = U3
ORF U3
T L R V V V R M P P Q H I K G D L K Q Q E F F L G C S K V S G V V
TGACTGGAAGATGCTGGATGAAGCTGTTTCCAAGTGTTCAGGACTATATTTCTAAATGGACCCAGCCTCTACCTGGGACTAAGCACTGAGTCCATC
ACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCCTGATATAAAGATTTTACC TGGGTCGGAGATGGGACCTTGATTCGTGACTCAGGTAG
pLM5 insert = U3
ORF U3
D V K M L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I

Tuesday, 18 November 1997 13:58
(fig 52 pLM5 (1 > 5425) Site and Sequence

Page 3

```
CATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCA
GTACCGATGTCGTAGTCGGTGCACTTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAG
pLM5 insert = U3
ORF U3
H G Y S I S H V K R V L D A E P P E M P P C R R G V N N I S V S L
AASGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCAAGCCGATGTCAGCACTACATAAGCCTCCTGCTGAAGCACCGGGC
TTCAGACTTCCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGGGTTCCGGCTACTACGTCGTGATGTATTTCGGAGGACGACTTCGTGGCCGC
pLM5 insert = U3
ORF U3
K G L K E K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R
CCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGC
GGAGCAGGAGAGCCCGGGTCCGCGTGCCCGTCTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACCTCGCGAGACCGGCACCTCCAGTGTCTCCCS
pLM5 insert = U3
ORF U3
L V L S G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G
ATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGATAGACCGGGAACAGGAATTGGGG
TAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCG
pLM5 insert = U3
ORF U3
I V S T F N M H Q Q S C K D L Q L Y L S N L A N O I D R E T G I G
ATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCTA
TACACGGGAGCCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGAGTGGACGTTTCATAGTATTACAGGGA
pLM5 insert = U3
ORF U3
D V P L V I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y
TATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCC
ATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTGGTACCGAAGTGAACTCGAAGTCCTACAAC TGGAAAGAGGTTGTTGCACCTCGGTCGG
pLM5 insert = U3
ORF U3
I I G T T N Q P V K M T P N H G L H L S F R M L T F S N H V E P A
```

Tuesday, 18 November 1997 13:56
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 4

AATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACGAAGAGCTGCTTCGGGTGCTCGACTGGGTAC
TTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCGTGCGTGTAGTTACGGTTGTTCCCTTCGACGAAGCCCACGAGCTGACCCATG 230

pLM5 insert = U3

ORF U3

N G F L V R Y L R R K L V E S D S D I N A N K E E L L R V L D V V

CCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTCTGTCGTGTCCTATGGCAATTGA
GGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAAC 240

pLM5 insert = U3

ORF U3

P K L V Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E

GGACTTCGGACCTGGTTCATTGACCTGTGGAACAACCTCTATCATTCCCTATCTACAGGAAGGAGCCAAAGGATGGGATAAAGTCCATGGACAGAAAGCT
CCTGAAGGCC TGGACCAAGTAAC TGGACACCTTGTGAGATAGTAAGGGATAGATGTCCTTCCTCGGTTCCTACCTATTTCAGGTACCTGCTTTTCGA 250

pLM5 insert = U3

ORF U3

D F R T V F I D L V N N S I I P Y L Q E G A K D G I K V H G Q K A

GCTTGGGAGGACCCAGTGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCACCCACCGTGG
CGAACCTCCTGGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGTGTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACC 260

pLM5 insert = U3

ORF U3

A V E D P V E V V R D T L P V P S A Q Q D Q S K L Y H L P P P T V

GCCCTCAGCAATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAACCT
CGGGAGTGTGTAACGGAGTGGAGGGCTCCTATCCTGTCAGTTTCTGTCGTGGGGTTCAAGAGACC TGAGTCTAGGAGAC TACCGGTACGACGACTTTGA 270

pLM5 insert = U3

ORF U3

G P H S I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L

TCAAGAAGCTGCCAACACATTGAGTCTCCAGATCGAGAAACCATCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCAACCC
AGTTCTTCGACGGTTGATGTAAC TCAAGGTCTAGCTCTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAATTCCTCAAGCCGTTAGTGACAGTGGGG 280

pLM5 insert = U3

ORF U3

O E A A N Y I E S P D R E T I L O P N L O A T L . G F G N H C H P

Tuesday, 18 November 1997 13:56
lig 52 pLM5 (1 > 5425) Site and Sequence

Page 5

CGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCCTCCTCTCCCTCTCTCTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAG
GCCGTGCTGCTTTCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGCGGGTCTCTCTTGTCTCTCTCT
pLM5 insert = U3
R T A E R V H Q L S . L L L S P L L F Q S T G S P A P G G E Q E G
GAGGAGATGAAGAGGAGGAGGACAGGTTCTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTTGTGCCCCCTAA
CTCTCTACTTTCTCTCTCTCTGTCGAAGAACCACGACATGGAAGCTCTTGAAGGATCTTCTTACCACCCACCGCAAAACCTTGAACACGGGGGATT
pLM5 insert = U3
G G D E R G G T G S V C C T F E N F L G R N G G V A F G N L C P L N
CACATTTACTGGCTCTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCCCTCCCCGT
GTGTAAATGACCGAGGAGATCTCCCGGCAAAATTTGGCGGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGGCA
pLM5 insert = U3
T F T G L L . R A R L N P L I S L D C A F . L P A I C C L P L P R
GCCTTCTTGACCTGGAAGGTGCCACTCCCACTGCTCTTCTTAATAAAATGAGGAAATGTCATCGCATTGCTGAGTAGGTGTCATTCTATTCTGGGG
CGGAAGGAACGGGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTTACTCTTTAAGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCC
A F L D P G R C H S H C P F L I K . G N C I A L S E . V S F Y S G
GGTGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGATGCTGGGGATGCGGTGGGCTCTATGGCTTCGAGGCGGAAAGAACCA
CCACCCACCCCGCTCTGTGCTTCCCCCTCTTAACCTTCTGTTATCGTTCGTCAGCCCTACGCCACCCGAGATACCGAAGACTCCGCCCTTTCTTGGT
G V G G A G Q Q G G G L G R O . O A C V G C G G L Y G F . G G K N Q
GCTGGGCTCTAGGGGTATCCACGCGCCCTGTAGCGGCGCATTAAAGCGGGGGGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGG
CGACCCGAGATCCCCATAGGGGTGCGGGGACATCGCGGTAATTCGCGCCGCCACACCAATGCGCGTCGCACTGCGATGTGAACGGTGGG
L G L . G V S P R A L . R R I K R G G C G G Y A Q R D R Y T C Q R
CCTAGCGCCCGCTCTTTCTGCTTCTTCCCTTCTTCTCGCCACGTTTCGCGGCTTTCCCGCTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGA
GGATCGCGGGCGAGGAAGCGAAGGAAGGAAGGAAGAGCGGTGCAAGCGGCCGAAAGGGGAGTTCGAGATTAGCCCCGTAGGGAAATCCCAGGGGT
P S A R S F R F L P F L S R H V R R L S P S S S K S G H P F R V F
TTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGAGGT
AAATCAGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTGCATCACCCTGAGCGGACTATCTGCCAAAAAGCGGAACTGCA
I . C F T A P R P Q K T . L G . V F T . V A I A L I D G F S P F D V
TGGAGTCCACGTCTTTAATAGTGGACTCTTGTTCAAAACCTGGAACAACACTCAACCTATCTCGGTCTATCTTTTGATTATAAGGGATTTTGGGGAT
ACCTCAGGTGCAAGAAATATCACTGAGAACAAGGTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAATAAATATTCCTTAAACCCCTA
G V H V L . . V T L V P N V N N T O P Y L G L F F . F I R D F G D
TTCGCCCTATTGGTTAAAAATGAGCTGATTTAACAATAATTAACGCGAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGC
AAGCGGATAACCAATTTTACTCGACTAAATTTGTTTAAATTCGCTTAATTAAGACACCTTACACACAGTCAATCCCAACCTTTTCAGGGTCCGA
F G L L V K K . A D L T K I . R E L I L V N V C O L G C G K S P G

Tuesday, 18 November 1997 13:58
fig 52 pLM5 (1 > 5425) Site and Sequence

Page 6

CCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
GGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTGGTCCACACCTTTTCAGGGGTCCGAGGGGTCTGCTTCATACGTTTCGT 3300
S P G R Q K Y A K H A S O L V S N Q V V K V P R L P S R Q K Y A K H
TGCACTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGCCCAGTTCGCCCATTCGCCCCCATGGCTGACT
ACGTAGAGTTAATCAGTCGTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGA 3400
A S O L V S N H S P A P N S A H P A P N S A Q F R P F S A P V L T
AATTTTTTTATTTATGCAAGGCGGAGGCCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTTAGGCTTTTGCAAAAAG
TTAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACTCTCCGAAAAACCTCCGGATCCGAAAAAGTTTTTC 3500
N F F Y L C R G R G R L C L . A I P E V V R R L F V R P R L L Q K
CTCCCGGAGCTTGATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGATGATTGAACAAGATGGATTGCACGAGGTTCTCCGG
GAGGGCCCTCGAACATATAGGTAAAGCCTAGACTAGTTCTCTGTCTTACTCTAGCAAGCGTACTAACTTGTTCTACCTAACGTGCGTCCAAGAGGCC 3600
A P G S L Y I H F R I . S R D R M R I V S H D . T R V I A R R F S G
CCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTGACGCGAGGGCGCCCGGT
GGCGAACCCACTCTCCGATAAGCCGATATGACCCGTGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTTCGCTCCCGCGGGCCA 3700
R L G G E A I R L . L G T T D N R L L . C R R V P A V S A G A P G
TCTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAATGCAGGACGAGGCAGCGCGGCATCGTGGCTGGCCACGACGGCGTTCCTTGCGCAGCT
AGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCTGCTCCGTGCGCGCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCTCGA 3800
S F C Q D R P V R C P E . T A G R G S A A I V A G H D G R S L R S
GTGCTCGACGTTGTACTGAAGCGGGAAGGGAAGTGGCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCCTGTCTCTCACCTTGCTCTCCGAGAAAAG
CACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTCACGGCCCCGTCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTT 3900
C A R R C H . S G K G L A A I G R S A G A G S P V I S P C S C R E S
TATCCATCATGGCTGATGCAATCGGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTAC
ATAGGTAGTACCGACTACGTTACGCCGCCGACGATGCGAACTAGGCCGATGACGGGTAAAGCTGGTGGTTCGCTTTGTAGCGTAGCTCGCTCGTGCATG 4000
I H H G . C N A A A A Y A . S G Y L P I R P P S E T S H R A S T Y
TCGGATGGAAGCGGCTTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTCGCCAGGCTCAAGGCGCGCATGCC
AGCTACCTTCGGCCAGAACAGCTAGTCTCTAGACCTGCTTCTCGTAGTCCCCGAGCGCGGTGCGCTTGACAAGCGGTCCGAGTTCCGCGCGTACGGG 4100
S D G S R S C R S G . S G R R A S G A R A S R T V R Q A Q G A H A
GACGGCGAGGATCTCGTCTGACCATGGCGATGCCCTGCTTGCCGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGG
CTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAATACCTTTTACCGCGGAAAAGACCTAAGTAGCTGACACCGGCGGACCC 4200
R R R G S R R D P W R C L L A E Y H G G K V P L F V I H R L V P A G
GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGTCTTACGGTAT
CACACCGCTTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCGCTTACCCGACTGGCGAAGGAGCACGAAATGCTATG 4300
C G G P L S G H S V G Y P . Y C . R A V R R M G . P L P R A L F Y
CGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCGCA
GGGCGAGGGCTAAGCGTCCGCTAGCGGAAGATAGCGGAAGAACTGCTCAAGAGACTCCCGCTGAGACCCCAAGCTTACTGGCTGGTTGGCTCCGGT 4400
R R S R F A A H R L L S P S . R V L L S G T L G F E M T D Q A I F

Tuesday, 18 November 1997 13:58
fig 52 pLMS (1 > 5425) Site and Sequence

Page 7

ACCTGCCATCACGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGS
TGGACGGTAGTGCTCTAAAGCTAAGGTGGCGCGGAAGATACTTTCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCC 510X
N L P S R D F D S T A A F Y E R L G F G I V F R D A G W M I L Q R G
GGATCTCATGCTGGAGTTCCTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAGCAATAGCATCACAAATTCACAAATAAGCA
CCTAGAGTACGACC TCAAGAAGCGGGTGGGGTTGAACAAATAACGTGGAATATTACCAATGTTTATTCGTTATCGTAGTGTTTAAAGTGTTTATTCGT 520X
D L M L E F F A H P N L F I A A Y N G Y K . S N S I T N F T N K A
TTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
AAAAAAGTGACGTAAGATCAACACCAAAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCCTATTAGTA 530X
F F S L H S S C G L S K L I N V S Y H V C I P S T S S . S L A . S
GGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGT
CCAGTATCGACAAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTGTATGCTCGGCCCTTCGTATTTACATTTTCGGACCCACGGATTACTCA 540X
V S . L F P V . N C Y P L T I P H N I R A G S I K C K A V G A . . V
GAGCTAACTCACATTAATTGCGTTG 5425
CTCGATTGAGTGTAATTAACGCAAC
S . L T L I A L

Tuesday, 18 November 1997 13:58

fig 53 pLM6 (1 > 4947) Site and Sequence

Enzymes : All 146 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

```
GTGGCACTTTTCGGGAAATGTGCGCGGAACCCCTATTGTATTCTTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAA~
CACCGTGAAAAGCCCCTTTACACGCGCCTTGGGGATAAACAATAAAAAAGATTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTA 100
V H F S G K C A R N P Y L F I F L N T F K Y V S A H E T I T L I H
GCTTCAATAATATTGAAAAAGGAGATGAGTATTCAACATTTCCGTGTCGCCCTTATCCCTTTTTTTCGGGCATTTTGCTTCTGTTTTGCTCAAC
CGAAGTTATTATAACTTTTCTTCTCATACTCATAAGTTGTAAAGGCACAGCGGGAATAAGGGAAAAAACGCCGTAAACGGAAGGACAAAAACGAGTS 200
A S I I L K K E E Y E Y S T F P C R P Y S L F C G I L P S C F C S
CCAGAAACGCTGGTGAAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTACATCGAAC TGGATCTCAACAGCGGTAAAGATCTTGAGAGTT
GGTCTTTGCGACCACTTTCAATTTCTACGACTTCTAGTCAACCCACGTCACCCAATGTAGCTTGACCTAGAGTTGTCGCCATTCTAGGAACCTCAA 300
P R N A G E S K R C . R S V G C T S G L H R T G S Q Q R . D P . E F
TTCGCCCCGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGCAAGAGCAACTCGGTCS
AAGCGGGGCTTCTTGCAAAAGGTTACTACTCGTGAATAATTCAAGACGATACACCGGCCATAATAGGGCATAACTGCGGCCCTTCTCGTTGAGCCAGC 400
S P R R T F S N D E H F . S S A M V R G I I P Y . R R A R A T R S
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTCTGCC
GGCGTATGTGATAAGAGCTTACTGAACCAACTCATGAGTGGTCAGTGTCTTTTCGTAGAATGCCACCGTACTGTCAATCTCTTAATACGTCACGACGS 500
P H T L F S E . L G . V L T S H R K A S Y G V H D S K R I M Q C C
ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCCTTTTTTGCAACAATGGGGGATCATGTAA
TATTGGTACTCACTATTGTGACGCCGTTGAATGAAGACTGTTGCTAGCCCTCTGGCTTCCTCGATTGGCGAAAAACGTTGTACCCCTTAGTACATT 600
H N H E . . H C G Q L T S D N D R R T E G A N R F F A Q H G G S C H
CTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCGTAGCAATGGCAACAACGTTGCGCAAACT
GAGCGGAAC TAGCAACCTTGGCTCGACTTACTTCGGTATGGTTTGTGCTCGCACTGTGGTGTCTACGGACATCGTTACCGTTGTTGCAACGCGTTTGA 700
S P . S L G T G A E . S H T K R R A . H H D A C S N G N N V A Q T
ATTAAC TGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAAGACTGGATGGAGGCGGATAAAGTTGCAGGACCAC TTCGCTCGGCCCTTCCG
TAATTGACCGCTTGATGAATGAGATCGAAGGGCGTTGTTAATTATCTGACCTACCTCCGCCATTTCACAGCTCTGGTGAAGACGCGAGCCGGGAAGGC 800
I N V R T T Y S S F P A T I N R L D G G G . S C R T T S A L G P S
GCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAE
CGACCGACCAATAACGACTATTTAGACCTCGGCCACTCGCACCCAGAGCGCCATAGTAACGTCGTGACCCCGGTC TACCATTGCGGAGGGCATAGCATC 900
G V L V Y C . . I V S R . A V V S R Y H C S T G A R V . A L P Y R S
TTATCTACACGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAAGTCTCAGACCA
AATAGATGTGCTGCCCTCAGTCCGTTGATACCTACTTGCTTTATCTGTCTAGCGACTCTATCCACGGAGTGACTAATTCGTAACCAATTGACAGTCTGGT 1000
Y L H D G E S G N Y G . T K . T D R . D R C L T D . A L V T V R P
AGTTTACTCATATATACTTTAGATTGATTTAAACTTCAATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCT
TCAAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAAACTATTAGAGTACTGTTTTAGGGA 1100
S L L I Y T L D . F K T S F L I . K D L G E D P F . . S H D O N F
TAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAAGAAAGATCAAGGATCTTCTTGAGATCCTTTTTTTCGCGGTAATCTGCTGCTTGCAAA
ATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTAGAAAGACTCTAGGAAAAAAGACGCGCATTAGACGACGAACTTT 1200
L T . V F V P L S V R P R R K D O R I F L R S F F S A R H L L . A H
```

Tuesday, 18 November 1997 13:58
fig 53 pLM8 (1 > 4947) Site and Sequence

Page 2

CAAAAAACCACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGCAGATACCAAA
GTTTTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCGATGGTTGAGAAAAAGGCTTCATTGACCGAAGTCGTCGCGTCTATGGTTT 1300
K K T T A T S G G L F A G S R A T N S F S E G N V L Q Q S A D T)
TACTGTCCTTCTAGTGTAGCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT
ATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGA 1400
Y C P S S V A V V R P P L Q E L C S T A Y I P R S A N P V T S G C
GCCAGTGGCGATAAGTCGTGCTTACCAGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGTTCTGTCACACAGC
CGGTCACCGCTATTGAGCAGAAATGGCCAACTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTCG 1500
C Q V R . V V S Y R V G L K T I V T G . G A A V G L N G G F V H T A
CCAGCTTGAGCGAACGACCTACACGAAGTGAATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGGGGACAGGTATCC
GGTCGAACCTCGCTTGGTGGATGTGGCTTGACTCTATGGATGTGCGACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTTCGCGCTGTCCATAGG 1600
O L G A N D L H R T E I P T A . A M R K R H A S R R E K G G Q V S
GGTAAGCGCAGGGTCGGAACAGGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCTGGTATCTTTATAGTCTGTGCGGGTTTCGCCACCTTGACTT
CCATTGCGCGTCCAGCCTTGCTCTCTCGCGTGTCCCTCGAAGGTCCTCTTTCGCGACCATAGAAATATCAGGACAGCCCAAGCGGTGGAGACTGAA 1700
G K R Q G R N R R A H E G A S R G K R L V S L . S C R V S P P L T
GAGCGTCGATTTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTG
CTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGGTGCTTTCGCGGAAAAATGCCAAGGACCGGAAAAACGACCGGAAAAAC 1800
A S I F V M L V R G A E P M E K R Q Q R G L F T V P G L L L A F C
CTCACATGTTCTTCTCGGTTATCCCCTGATCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAACGACCGAGCG
GAGTGTAAGAAAGGACGCAATAGGGGACTAAGACACCATTGGCATAATGGCGGAAACCTACTCGACTATGGCGAGCGGCGTCGGCTTGTGGCTCGC 1900
S H V L S C V I P . F C G . P Y Y R L . V S . Y R S P Q P N D R A
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACGCCCTCCTCCCGCGGTTGCGCGATTCTTAATGACGTGGCAGCAGAGTTT
GTCGCTCAGTCAGTCGCTCCTTCGCCCTTCGCGGGTTATGCGTTTGGCGGAGAGGGGCGCAACCGGCTAAGTAATTACGTCGACCGTGCTGTCCAAA 2000
O R V S E R G S G R A P N T Q T A S P R A L A D S L M Q L A R Q V
CCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGT
GGGCTGACCTTTCCGCCGCTACTCGGTTGCGTTAATTACACTCAATCGAGTGAGTAATCCGTGGGGTCCGAAATGTGAAATACGAAGGCCGAGCATACA 2100
S R L E S G Q . A O R N . C E L A H S L G T P G F T L Y A S G S Y V
TGTGTGAATTGTGAGCGGATAACAATTTACACAGGAACAGCTATGACCATGATTACGCCAAGCGGCAATTAACCTTACATAAGGGAACAAAAGCT
ACACACCTTAACACTCGCCTATTGTTAAAGTGTGCTTTGTGATACCTGGTACTAATCGGTTTCGCGGTTAATTGGGAGTGATTTCCCTTGTTCGTA 2200
V V N C E R I T I S H R K Q L . P . L R Q A R N . P S L K G T Y A
GGGTACCGGGCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCGAATTCTTCAGCCCGGGGATCCATGCAATAGGAGGAGGAGCCAGAGA 2300
CCCATGGCCCGGGGGGAGCTCCAGCTGCCATAGCTATTGCAACTATAGCTTAAGGACGTCGGGCCCCCTAGGTACGTTTACTCTCTCTCTCGGCTCTCT
G Y R A P P R G R R Y R . A . Y R I P A A R G I H A N E E E E P E
U3 stuk

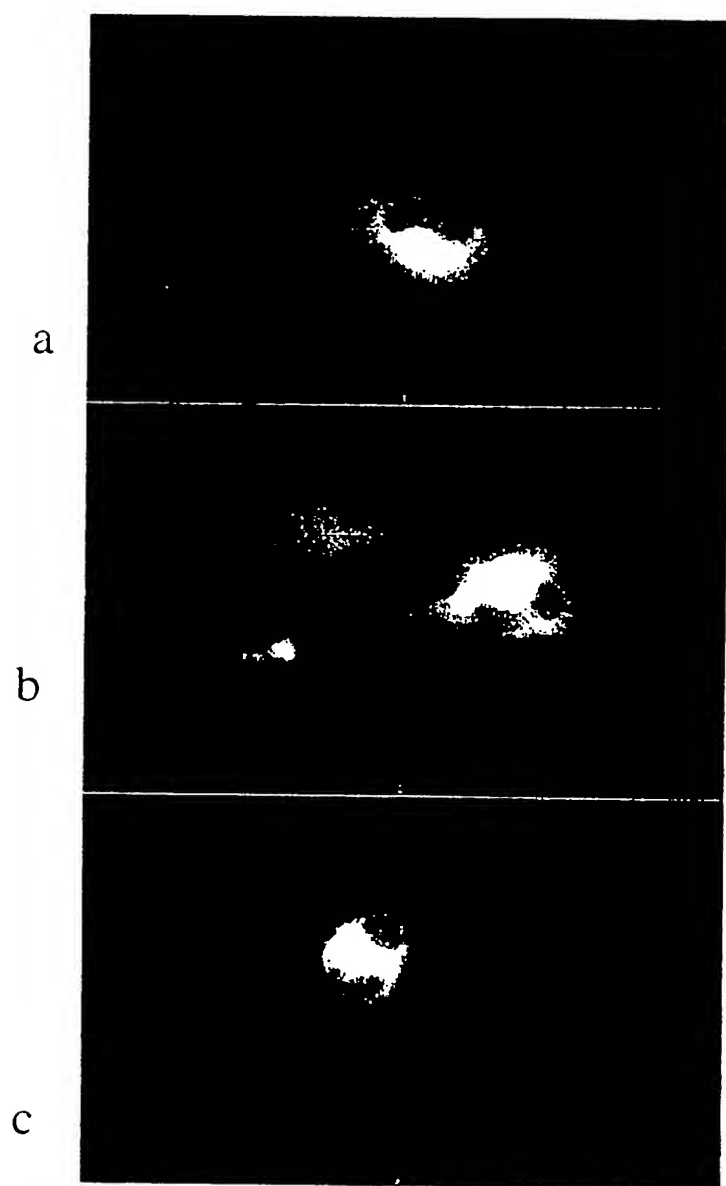


Figure 43

donderdag, 27 november 1997 15:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence
Enzymes : All 148 enzymes (No Filter)
Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

ATGACCATGATTACGCCAAGCTTGCAATGCTGCGAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAGAAATTGGAGCAACTACC 100
CACATCCATTATGCCACCCGCGGTTTCTAAGTGAGTTAATTTTGAGTTTACGACTACAAAATGTGTTCTTTAATACTATCTTCGACTTGAGTCTATT 200
CTGTATGACTAGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAAGGAACATTATTTACTTTGCTTATTTGCCCTAATTTGATTTAGTTTTCGATC 300
AACTAGATCTTACAAAATTGCAATACAAATTCATTTTCAGATTACCTCGCCACGTGTCGCCACGTGAGCAACCGCTTCAGCAACTAACCCAAATTCGA 400
ACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGATATCGAAAATTTGGTAAGAATTTTATTTTGAGCTCAAACTGTATAAAATGCC 500
CAGAAAAGAGATGATAAAATGTAGTTTTCGAAAACCTCCACCTTTATGCTCTAATATGACGGCTTATCTCAATTTTCTTGAGTTTATCAAA 600
AAATTTTCCACTATACAAATGTAGAAAAGTATTTTGACAAATTTGTGAGTTGACAGCTTTGTAAATAGATCCAAATGGAACCTAGATACAAGCTGTAA 700
AGTGAAGGAGCGCAAGCTATATCGGAAATAATGATCTGAAACAAATTTGTCTATTCTCAATGTTTAAAGACATGTTTGAAGATTTTCAAAATCG 800
CACTAGTTTCAGAACCTTCTTTTGTATGAAAAGTAAAAAACTATTTCAAACTCCACGCCACCATGTTTCACTCTTAATTTTATAAAAATTT 900
TGCAATTTACAAATCGCTCCCTTTGTCGCGAAAAGTCCCAACCAATCAATTTCTCGGCTTCATAATGACTTTTAAATGATGTGAGAAAACACAGAAG 1000
AGGCTAACTAAATTTGACAGGGACAGGTGTGCTCTTCT 1100
TTGTCATTTTCTTATAAACTTTGTGTGTGGAAGGAACTACACGGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTAATCTTTTCGGAAT 1200
TGATGAATAAGATAGAGCCGATGACACTGGCTGGTAG 1300
CATCACTGACAATTAATGTGCGGTTTATGCGCTCTTTCTTATCCGCCACTCATTCTGGGTTACCAAACTGGAATACATTTTACTACTATTCAGCC 1400
ATTTATTTTGATATTTAATTTTGTGCAATTAGGATATAACGACCTTTTAAAGTTTATTTAAAAAACGATATTTTCGATTTTAAAAATCTGAAAGT 1500
TTCAAAAATCAATAAATATTCCTAACAAATTTGTATGGCTAAATTTTATTTCTACTGTTGACAATATCTTTATATGTATCACTGTTTTCATCTCAAA 1600
ACCTTGAATCCCCAAGTTATAGGAAGCTCCGTGTACATTTCCCATGCTAATGATCGTACTCAGCACATATCCAAAAATTAAGCTAGACGGTTGATAA 1700
TTATTGGCAGCGTAATAAAGTGCAAGCAGTTAGAATTTAATTCAAGCACAGATTATCTATCAAAATCAATCTTTGAACATTGAGCCAGTTCGTACAA 1800
TTTTCTATGCTTTTGGCCATTAAAAAATTTCTCACCTCTTCATCCATCTCACTCGTATCAAAAAATATAGCAAAAGCCCGACTCTACTTTTAAAG 1900
AGAAGGAGATACTGAGCCACATGGCGTGTGACCTTTTCATCTCGTCTGCTCGGTCTCAATTCACGCTCATCTAATCTTCAATAGCCATAGACTC 2000
CTTGTTTCTTCTCGTTTGACTCGCGCTATTTTGTGGCTGCTGAAAGCCGGGAAATTTAGTATATTTAGAGCTTATCTTTATGCAATACATA 2100
AAAAACGAGGCAATTTAAAAATATTAATTAATGAGGTTGTAGATGTAGATTGGAAAAGAGAAAAAACAACAAATAGGAACCGCCAGATCAAAA 2200
TTCTATTTAAAGGTTTCAAGATGTTTAGGCAAGATTGGCTGAACAGAAAATGAAGTGCCTGCATAAATCTAGTGTAACTTTAGATTGAACCTCGGAA 2300
ATCTAAGCCTGAACATATAGCTTATTTAGATCTTAGTTGCGCATAGCTCAAGCCCAAGCAGAAATGACTTGCAATTTAGTTTAAAGCTAGATTGACT 2400
GCTTGCTTCAGTCTAATCCAGACTAGATTTCCAGAGAGTTTCAATTTTAAATGTTTCCAGTTTCTTGTACTTAAATCTTAATGCCCTGTGATGCGT 2500
AAAACTGTTATCCCTTCTCTCACTTTCAATTACAGATTCAATCAAGATTGGTATCAAGCCAAAGACGTCTGGACTTAAACCCCTCATCATCAACC 2600
ACTTCATCAAAATACAAATTCATTCGCTCGTCCAGCGTTCGAGTGGCAATAAATGTTGGCTCGACGATATCCACATCTCGCAAGAGCTTACGTA 2700
TCCGATCCCTTCGGCTTCTTTTGAAGATTTATTTATTTAGAAATCATCATCAAGTACAGCTCTATTTGAAATCTAAACCGACTACCTCCCACTCCA 2800
AAAACTCTTAGACCCACAAACCCAGCTAGTTGCTGTTGCTACAACTACAAAAATCGGAAGCTCAAGCTAGCGCTCTCGAAGGCTGAGCACCCCAAAA 2900
CTTGCTTCTGTGAAGACTATTGGAGCAAAAAGAGCCCGATAACAGCGGTGCTGTTGGTGGTGAATGCTGAAATTAAGTTATTAGTAGCAAAAACC 3000

dondorag, 27 november 1997 16:48
Fig 44 pNP9 Map (1 > 10122) Shs and Sequence

Page 2

CATCTTCTCATCGAATAGCCCAACCTACGAGAAAGCGCGCGGGTCCCTCAACAACAACTTTGTGAAAATCGCTGCCCAAGTGAAGAGTGGCT 3100
GAAGCGCGGACAGTAAGCTGGGAAGTCCACGTCTATGTGGAAGCTTTGTACGCCAAAAGTTTCTACCGTAAACCGGACGCCCCAATCATATCTCAA 3200
CAAGACTCGAAACGATGCTCAAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCAGCTGCCAACGTCAATCGAGCGGAAGGTTCCCTAA 3300
GCATGCATTCCATCTTCCAAGAGTTCAACGTGAGCGAAAAGTCTCGTCATCAGACGATCTTACTCTTAACGCCCTCCATCGTGACAGCTATCAGACA 3400
GCGGATAGCGCAACACCGGTTTCTCAAATATTATCAACAAGCCTGTTGAGGAAAAACCAACTGGCAGTGAAAGGAGTGAAGACAGCGAAAAA 3500
GATCCAGCTCAGCTGTTCCGCCAGTGACACCCAGCCAAATCGGAGTTGTTAGTCCAAATTATGGCACAATAAGAGTTGACAAATGACCCCGTGATAT 3600
CTGAAAACCGAGACCTGAAAAGCTCCAATCAATGAGCATCGACAGCGGAGCTTCCACGCTTCCACTCTAAAATCAGTTGTTCCACTTAAAATGAC 3700
TTCAATCCGACAACCAACACGTACGATGTTCTTCTAAACAAGGAAAAATCAGTCGCCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCCGCTGTGAA 3800
GACTCCATTGTGGCTCATGCGTGGCTCAGGTGACTCCGCCGCAAAAATCTTGGTAATCATTGCTGGAGAGAGGATGGGAAAGAAATAGCATCAG 3900
AATCAGCGGCTACACTCTGACCGCGGTGTTGCGATGTGCGCAAAATGAGGGAGAGCTGAAAAGATACGATGACATGACTGCTGAGCACAGAACGG 4000
CTATCTGACAACTCGAAGACAGTTCTCTTGTCTGTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGAGTAGAC 4100
ATGGCAACAGTGGCTCCAAACATAGCGACTATTCCCACTTTGTCGCCATCCACGCTTCTTCTCTCAAGCCCGAGTCCCACTCGGTCTCCACAT 4200
CAGTCGATTCTGATCTGAGCAGAACAGGAGATGTGTACAACTTCTGTCCAGTGCCGACGAGCCACGTCGGCGCGCTGCCACCTCAACCTTCGG 4300
ACAACATTGCTAAGATCCCCGGGATACTCATCTATTCTCCACACTTATCAGTGTGAGTGATAAGGACACAATGTCTATGCACTCAGAGACTAGTCGA 4400
CGAECTTCTTCAAAAACCAAGCTATTGAGGCAATTTCACTCACTTGATCGTAAATGCCACCTTCAAGAGTTCACATECCCGAGCAGCAATGGCGG 4500
CTCTTTGAGCGCGAGCGGGTCCGAACTCGATGTGAAATATGATTCTTCAGGATECTACTCGCGCGTTCCCGAGGTGGAAGCTCTACTGGTATCTA 4600
TGGAGAGAGCTTCAACTGCACAGACTATCCGATGAAAAATCCCCCGCAATTCTGCAAAAGTGAGATGGGATCCCACTATCACTGGTAGCAGACA 4700
GCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCGGACATGGCAGTGACTTGGAGTGTACAAGAACACTGTGACTCACTAACCAAGAAC 4800
AGGAGAACTATGGAGCAATTGTTGATCTTTTGGCAAAAGCTTAGAAAACCTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACG 4900
ATTGAGCGAGGACATTGCTCAATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGTAAAGAGGCGCTGGTGAGCTTCTTGTCAACCA 5000
TCTCTGGAATCAGTTGCATCCCATCGATCATGTGTCATCGTGTGAAAAAGCAGCAAGCAGGAGAGATCAGCTTGAGCTGTTTGGCAAGAACAGA 5100
AGAGCTGGATCCGCTCCTCACTCTCCAAGTTCAACAAGAGAGAACAAAGAACTACGACGAAGCACAATATGCCATCAATTTCCGGATCTCAAGGAACTCT 5200
TGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAGAACGCGATAGTGCACTTTACGAAGTCCGCTTGACAATCTGGATGTCGCCCGGAGTT 5300
GATGTTCTGAGGAGACAGTGAACAAGTTGAAAAACGAGAACAAAGCAATTAAGAAAGAGTGGACAACTCACCAACGGTCCAGCCACTCGTGTCTT 5400
CCCGCGCTCAATCCAGTTATCTACGACGATGAGCATGTCTATGATGAGCGGTGAGCAGTACATCAGCTAGTCAATCTCGAAACGATCTCTGGCTG 5500
CAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGGAAATCAGTTGATCGTTAAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCA 5600
ACCAGTCAGTCATGCTGGAAGACATTGATGTTCTATTTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGAGCATCAACTGGAAATCGATG 5700
CTCGTGATTCTATCTTGGCTATCAAAATGGTGAACCTCGACGCGTCACTGGAGCTCCACAACCATGATAACCAGCCATCAACTGACATTTACTTC 5800
CTCACTACAATCGAATGTTATGCACGGTCCGCCAGAGTCCGCTAGACAGTCTGGTCTTGATATGCTTCTTCAAGCAAAATGATTCTCAACTC 5900
GTCAAGTCAATTTGACAGAGAGAGCTCTGGTGTAGCTGGAGCACTGGAAATGGAAAGACAACTGGCGAAGACCTGGCTGCTTATGTATCTATTC 6000

donderdag, 27 november 1997 18:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence

Page 3

GAACAAATCAATCCGAGATAGTATTGTTAATATCAGCATTCCTGAAAACAATAAGAAGAAATTGCTTCAAGTGGAAACGACGCTGGAAAAGATCTTGAG 6180
AAGCAAGAATCATGCATCGTAATCTAGATAATATCCAAAGAAATCGAATTGCAATTTGTGTATCCGTTTTCGAAATGTCCACTTCAAAAACACGAA 6280
GGTCCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCAAATTCACCACAATTTCAAATGTCAAGTAATGTGGAATCGTCTCGAAGGAT 6380
TCATCTACGTTACCTCCGACGACGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTTTCAAAATCATTGACTTCTTCCCAAT 6480
AGCTCTTCAGGCCGTCAATAATTTTATYGAGAAAACGAATTCGTGTGTGTGACAGTTGGTCCAAGAGCATGCTTGAAGTGTCTCTAACTGTCTGATGGA 6580
TCCCGTGAATGGTTCATTGATTTGGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCTGCTCACTTCT 6680
TCGAGGATCCACCGACATCGTCTCAAAAAATGGCCGTGGTTCATGGTGAAACCCGGAATGTGCTCAACGCTTCACTCCAAAGACCTCGTCCC 6780
GTCACTGCCAACTCATCCGACAACTTCAATCCCTCGAGTGGTGTCAATTGCAATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAA 6880
GACTCTAATCTTCTCTCGCTCTCCCGCTTCTCTTATCTTCTGACCGGTACCATGGTATTGATATCTGAGCTCCGCATCGGCGGTGTCTATCAGATCG 6980
CCATCTCGCGCCGTGCTCTGACTTCTAAGTCAATTAATCTTCAACATCCCTACATGCTCTTCTCTCTGCTCCACCCCTATTTTGTATTAT 7080
CAAAAAAATCTTCTTAATTTCTTTGTTTTAGCTTCTTTAAGTCACCTCTAAACATGAAATTTGTAGATTCAAAAATAGAATTAATTCGTAATAA 7180
AAAGTGAAAAAATTTGTGCTCCCTCCCCCAATTAATAATCTATCCAAATCTACACAATGTTCTGTGTACACTTCTTAATGTTTTTTTACTTCT 7280
GATAAATTTTTTTGAAACATCATAGAAAAACCGACACAAAATACCTTATCATATGTTACGTTTCAAGTTTATGACCGCAATTTTTATTTCTTCGCAG 7380
TCTGGGCTCTCATGACGTCAAATCATGCTCATCGTGAAGGTTTGGAGTATTTTGGAAATTTTCAATCAAGTGAAGTTTATGAAATTAATTTTCC 7480
TGCTTTGCTTTTGGGGCTTCCCTATTGTTGTCAAGAGTTTCGAGGACGGGCTTTTCTTGCTAAATCACAAGTATTGATGAGCAGATGCAAGA 7580
AAGATCGGAAGAAAGTTTGGGTTGAGGCTCACTGGAAGGTGAGTAGAATGATAATTTGAAAGTGGAGTGTCTATGGGCTTTTGCCTTAATGA 7680
CAGAATACATCCCAATATACCAACATAACTGTTTCTTACTAGTGGCGCTACGGGCCCTTCTGCTCGCGGTTTGGGTGATGACGGTGAACCTCT 7780
GACATATGAGTCTCCGAGACGGTCAAGTTGTGTGAAGGGATGCCGGAGCAGACAAGCCCTGTCAGGGCGGTCAGCGGCTGTTGGCGGGGTGTCG 7880
GGGCTGGCTTAATATGCGGCATCAGAGCAGATTGTACTGAGAGTGACCATATGCGGTGTGAAATACCGCAGATGCGTAAGGAGAAAAATACCGCATC 7980
AGCGGGCTTAAGCGCTGTGATACGCTATTTTATAGGTTAATGTGATGATAAATGGTTTCTTAGACGTGAGGTGGCACTTTTCGGGGAATGTG 8080
CGCGGAACCTTATTTGTTATTTTCTAAATACATTCAAATATGATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGA 8180
AGAGTATGAGTATCAACATTTCCGTGTCGCCCTTATCCCTTTTTTGGGCAATTTTCCCTTCTGTTTTTGTCTACCCAGAAACGCTGTGAAAGTAAA 8280
AGATGCTGAAGATCAGTTGGTGCAGAGTGGTTACATCGAAGTGTCTCAACAGCGGTAAGATCCTTGAGAGTTTTGCCCCGAAGAACGTTTTCCA 8380
ATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGGGTATTATCCGATTGACGCCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGAAATG 8480
ACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAAGTGTGCTAACCATGAGTGATAACACTGC 8580
GGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAACATGGGGGATCATGTAATCTCGCTTGTGCTGGGAACCG 8680
GAGCTGAATGAAGCCATACAAACGACGAGGTGACACCCAGTGCCTGTAGCAATGGCAACAGTTGCGCAACTATTAACTGGCGAACTACTTACTC 8780
TAGCTTCCCGCAACAAATTAATAGACTGGATGGAGGGGATAAGTTGCAAGACCACTTCTGCGCTCGGCTTCCGGTGGCTGGTTTATTGCTGATAA 8880
ATCTGGAGCCGGTGAAGCTGGGCTCTCGGGTATCTTCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGCAGGGGAGTCAAG 8980
GCAACTATGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAGCATTGGTAAGTGTGACCAAGTTTACTCATATATACTTTAGA 9080

donderdag, 27 november 1997 18:48
fig 44 pNP9 Map (1 > 10122) Site and Sequence

Page 4

TTGATTAAAACTTCATTTTAAATTTAAAAGGATCTAGGTGAAGATCCTTTTGATAATCTCATGACCAAATCCCTTAACGTGAGTTTCTTCCACTG 9100
AGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTCTGCGCGTAATCTGCTGCTTGCAAAACAAAAAAACACCCGCTACCAGCG 9200
GTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTTCCGAAGGTAACGGCTTCAGCAGAGCGCAGATACCAAACTACTGTCTTCTAGTGTAGCCGT 9300
AGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCT 9400
TACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGCTCGGGCTGAACGGGGGTTCTGTCACACAGCCAGCTTGGAGCGAACGACCTAC 9500
ACCGAACTGAGATACCTACAGCGTGAGCAATTGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGGCGACAGGTATCCGTAAGCGGCGAGGTCTGGAAACAG 9600
GAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCTGGTATCTTTATAGTCTGTGCGGTTTCCGCACCTCTGACTTGAGCGTCTGATTTTGTGATGCTC 9700
GTCAGGGGGGGGAGCTATGGAAAAACGCCAGCAACGCGGCCCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTGTCTCAGATGTTCTTCTCTCGTTA 9800
TCCCTGATTCTGTGGATAACCGTATTACCGCTTTTGAGTGAGCTGATACCGCTCGCCGAGCCGAACGACCGAGCGACCGAGTCACTGAGCGAGGAAG 9900
CGGAAGAGCGCCCAATACGCAACCGCTCTCCCCGCGGTGGCCGATTCTTAATGCAGCTGGCAGCAGGTTTCCCGACTGCAAGCGGGCAGTGA 10000
CGGCAACGCAATTAAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGAATTGTGAGCGGATAA 10100
CAATTCACACAGGAAACAGCT 10122

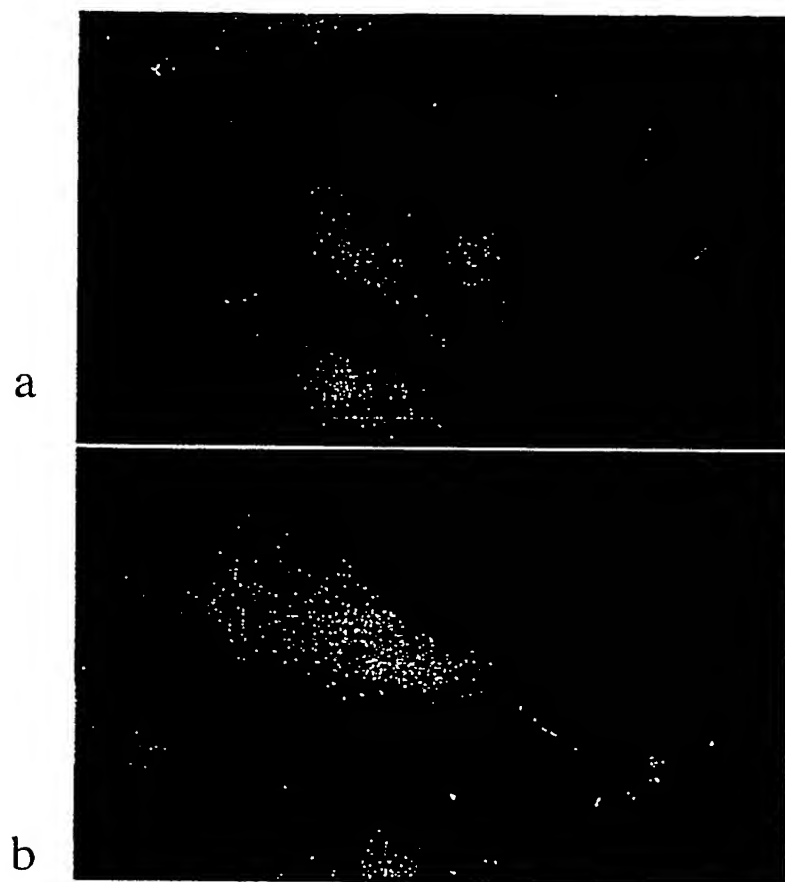


figure 45

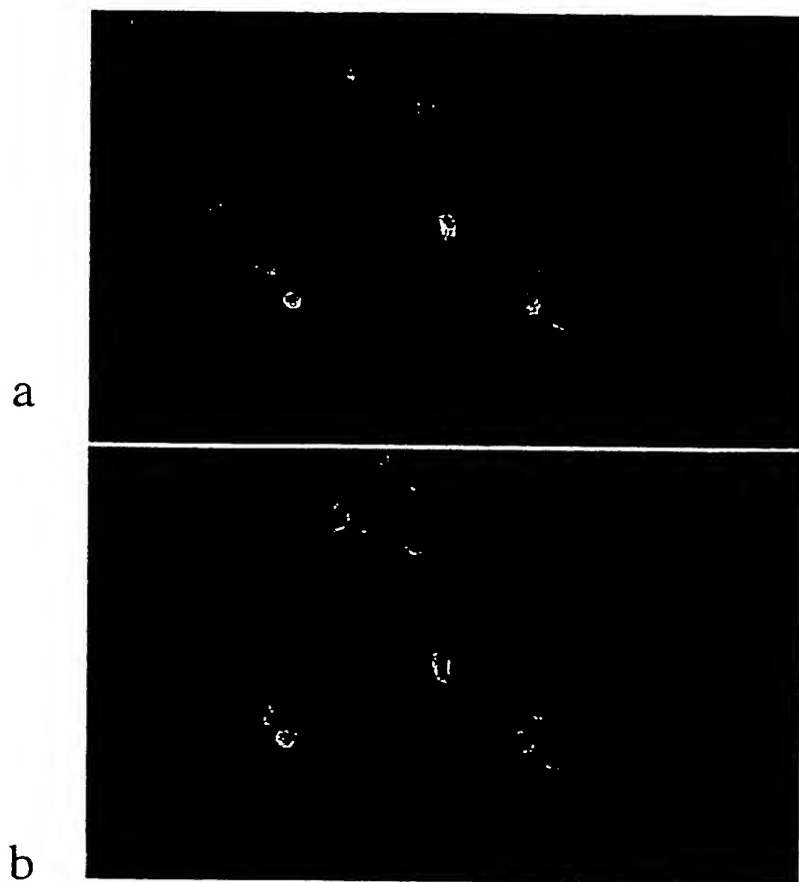


Figure 46

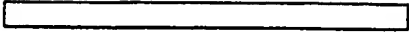










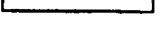
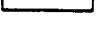
		+ end	MTB	actin
pTB72		+	+	+
pcDU2		nd	nd	+
pcDU3		nd	nd	+
pcDU4		nd	nd	+
peGFP72		+	+	nd
peGFPsma		- nd	+	
peGFPecf		-	+	nd
peGFPxba		-	-	nd
peGFPsac		-	-	nd
pLM4		+	+	nd
pCB251		nd	nd	nd
pLM5		nd	nd	nd
pCB201		nd	nd	+

Figure 47

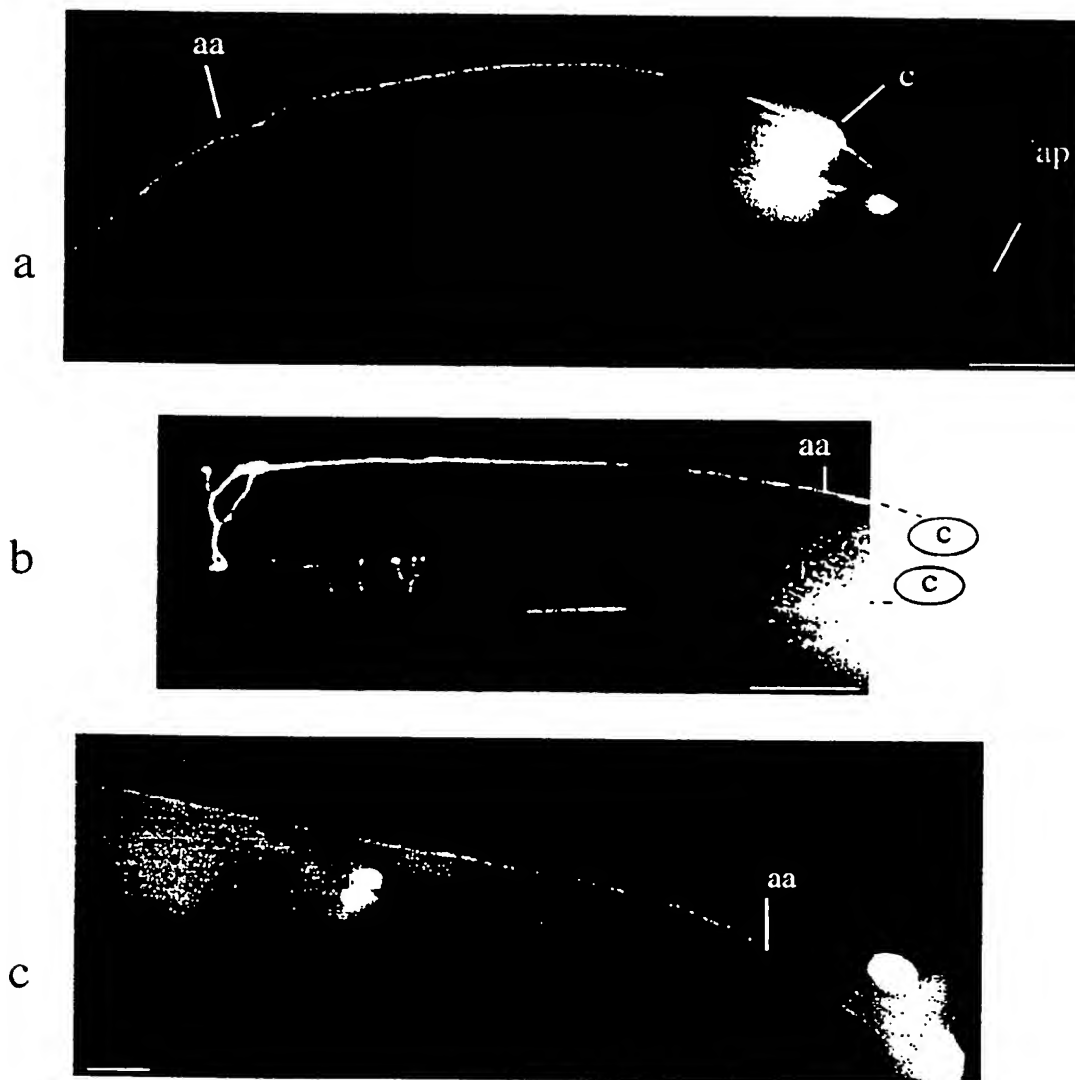


Figure 50

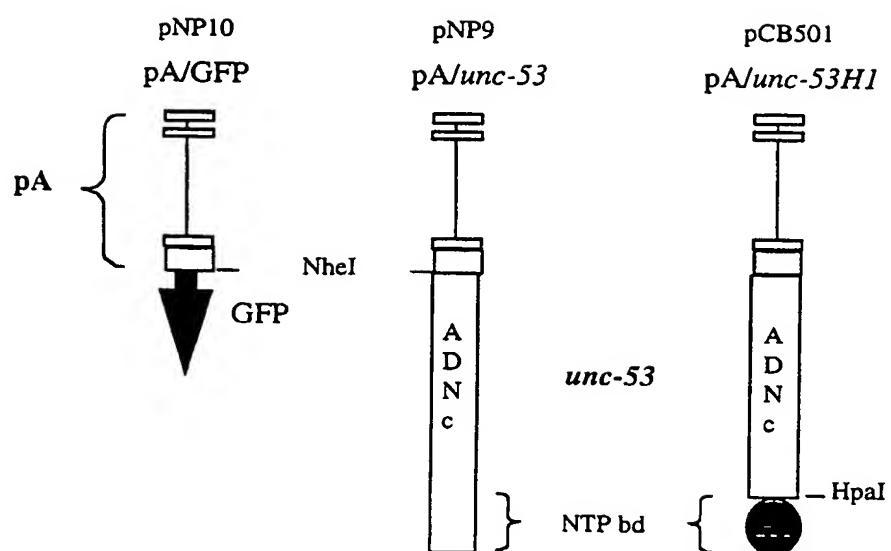


Figure 51a

Monday, 27 November 1997 16:48
fig 35 pNP9 Map (1 > 12641) Site and Sequence

Page 3

GATTTTCAGACTTGGGCTATAAATTTTGTCAAACCTAGGAATCTTAAATATTTGTATTTTTCGAAGAATGCTCTCAATCTCAAATTCATATTTATA 6100
ATTTTCAGACCCCGTGATATCTGAAAAACAGAACCTGAAAAGTCCAAATCAATGACATCGACACGACGGAGCTTCACCGCTTCCACTCTAAATACAG 6200
TTGTTCCACTTAAATGACTTCAATCCGACCAACCAACGTACGATGTTCTTCTAAACAAGGAAAAATCAGATCGCTGTCAAGTCTTTGCATATCA 6300
GCAGTCTCCGCTCTGAAGACTCCATTGTGGCTCATGGCTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTCGCTGGAGAGAGGATG 6400
GGAAAGAATAAGACATCAGAAATCCAGCGGTACACCTCTGACGCCGCTGTTGCGATGTGCGCCAAAATGAGGGAGAGCTGAAAGAATACGATGACATGA 6500
CTCGTCGAGCAGACAGCGGTATCTGACAACTTCAAGACAGTTCTCTCTGCTCTGGAATATCCGATAACACGAGCTCCAGACATATCCACGGA 6600
CGATTTGTCCGGAGTAGACATGGCAACAGTCGCTCCAAACATAGCGACTATTTCCACTTTGTTGCGCATCCACGCTCTTCTCTCAAGGCCCGGAGTC 6700
CCCAGTGGTCTCCACATCAGTCGATTCTGATCTCGAGCAGAACAGGAGAAATGTGTACAACTTCTGTCCAGTGGCGAACGACCAACGTGGCGCG 6800
CTGCCACTCAACCTTCGGACAACTTCCTAAGATCCCGGATACCTCATCTATTCTCCACACTTATCAGTGTGAGCTGATAGGACACAATGTCTAT 6900
GCACTCAGAGACTAGTCGACGACTTCTTCAAAAAACCAAGCTATTAGGCCAAATTCATTCACTTGATCGTAAATGCCACTTCAAGATTACATCC 7000
ACCGAGCAGACAAATGGCGGCTCTTGTAGCCCGAGACGGGTGCCAACTCGATGTGAAATATGATTCTTCAGGATCTACTCGGCGGTTCCCGAGGTG 7100
GAAGCTCTACTGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGCACTTCTGCCAAAAGTGAGATGGGATCCCACT 7200
ATCACTGGCTAGCAGCAGCATATGGATCTCTCAATGAGAAGTACGAACATGCTATTCCGGACATGGCAGTGACTTGGAGTGTACAGAACAATGTC 7300
GACTCACTAACCAAGAAACAGGAGAACTATGGAGCAATTTGTTGATCTTTTTCAGCAAAAGCTTAGAAAATCACTCAACACATTGATCGATCCAACTGA 7400
AGCCTGAAGAGGCAATACGATTCAGGAGGACATGCTCATTTGAGGGATAITAGCAATCATCTTGCATCCAACTCAGCTCATGTAACGAAGCGCGTGG 7500
TGAGCTTCTCTGCAACATCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCTGCAAAAGCAGCAAGCAGGAGAGATCAGCTTGAGC 7600
TCGTTTGGCAAGAACAGAGAGCTGGATCCGCTCTCACTCTCAAGTTCCACAGAGAGAGAACAGAACTACGACGAGCACAATATGCCATCAATTT 7700
CCGGATCTCAAGGAATCTTGACAACATGATGTGATTGAGTGAAGCAAGAGCTCAAGAACCGGATAGTGCACTTTACGAAGTCCGCTTGACAATCT 7800
GGATCGTGCCCGCAAGTTGATGTTCTGAGGAGACAGTGAAACAGTTGAAACCGAGAACAGCAATTAAGGAAGAGTGGCAAACTCACCAACGGT 7900
CCAGCCACTCGTGCTTCTCCCGCGCTCAATTCAGTTATCTACGACGATGAGCATGTCTATGATGCACCGTGATGAGTACATCAGCTAGTCAATCTT 8000
CGAAACGATCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTGATCGTTAACCCGGACAAAGAGATAATCTG 8100
AGGATATCTTGCATGTCAACCACTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGCTACCTATCCAGAAATGATGTGGAG 8200
CATCAACTTGGAAATCGATGCTGCTGATCTATCTCTGGCTATCAAAATGGTGAACTTCAGCGGTCATTGGACACTCCAAACCATGATAACCAAGCCATC 8300
CAACTGACATTTCTACTTCTCACTACAATCCGAATGTTATGACCGGTGCCGACAGAGTCGCTAGACAGTCTGCTCTTGATGATCTCTTCCAAA 8400
GCAATGATTTCTCAACTCGTCAAGTCAATTTTGACAGAGAGAGCTCTGGTGTTAGCTGGAGCAACTGGAATTGGAAAGAGCAAACTGGCGAAGACCTG 8500
GCTGCTTATGTATCTATTGGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCAATTCCTGAAAACAAATAAGAAAGAAATGCTTCAAGTGGCAACGAC 8600
GCTGGAAAGATCTTGAGAGCAAGAAATCATGCATGTAATCTAGATAATATCCAAAGAAATCGAATTGCAATTTGTGTATCCGTTTTTTCAAATGT 8700
CCCACTTCAAAACAAACGAAGTCCATTTGTAGTATGCACAGTCAACCGATATCAATCCCTGAGCTTCAAAATCCACAATTTCAAAATGTCAGTAATG 8800
TCGAATCTCTCGAAGGATTCATCTACGTTACCTCCGACGAGGGCGGTAGAGGATGATATCGTCTAAGTGTACAGATGCCATCAGAGCTCTTCAAAA 8900
TCATTGACTTCTTCCCAATAGCTCTTCAGGCGGTCAATAATTTTATTGAGAAAACGAATTCGTTGATGTGACAGTTGGTCCAGAGCATGCTTGAATG 9000

dunderdag. 27 november 1997 16:46
Rg 35 pNP8 Msp (1 > 12641) Site and Sequence

Page 4

TCCTCTAACTGTCGATGGATCCCGTGAATGGTTCAATTCGATTGTGGAAATGAGAACTTCATTCATATTTGGAACTGTTGCTAGAGATGGCAAAAAACC 9180
TTGGGTGCTGCACCTCTTCGAGGATCCACCGACATCGTCTCTAAAAATGGCGTGGTTCGATGGTGAACCCCGGAGAAATGTCTCAAACTGCTTC 9200
AACTCCAAGACTCTGCTCCGTCACCTGCCAACTCATCCGACAACACTTCAATCCCTCGAGTCTGTCATCCAAATTCATGCTACCAAGCATCAGACCAT 9300
CGACAACATTTGAACAGAAGACTCTAATCTTCTCTGCTCTCCCGCTTCTCTTATCTTCGTACCGGTACCATGGTATTGATATCTGAGCTCCGCATC 9400
GGCCGCTGTATCAGATCGCCATCTCGCGCCGCTGCTCTGACTTCTAAGTCCAATTAATCTTCAACATCCCTACATGCTCTTCTCCCTGCTGCCAC 9500
CCCTATTTTTGTTATTATCAAAAAAATCTTCTTAATTTCTTTGTTTTTAGCTTCTTTAAGTCACTCTAACAATGAAATTTGTGTAGATTCAAAA 9600
TAGAATTAATTCGTAATAAAAGTCGAAAAAATTTGTCTCCCTCCCCCATTAAATAAATCTATCCCAAAATCTACAAATGTTCTGTGTACATTC 9700
TTATGTTTTTTTACTCTGATAAAATTTTTTGAACATCATAGAAAAACCGCACAAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCA 9800
ATTTTTATTTCTCGCAGCTCTGGGCTCTCATGACGTCAAATCATGCTCATCGTGAAGGTTTTGGAGTATTTTTGGAATTTTCAATCAAGTGAAG 9900
TTTATGAAATTAATTTCTGCTTTTGGTTTTGGGGTTTCCCTATTTGTTGTCAAGAGTTTCGAGGACGGGTTTTTTCTGCTAAAATCACAAGTAT 10000
TGATGACGACGATGAAGAAAGATCGGAAGAAGTTTGGGTTTGAAGGTCAGTGGAGGTGAGTAGAAGTTGATAATTTGAAAGTGGAGTAGTGTCTATG 10100
GGGTTTTTGCCTAAATGACAGAATACATTCCTCAATATACCAAACTAATGTTTCTACTAGTCGGCGTACGGGCTCTTCTGCTCGCGCTTCCGT 10200
GATGACGGTGAAAACTCTGACACATGACGCTCCCGAGACGGTCAACGCTTGTGTGAAGCGGATCCGGGAGCAGACAAGCCCTCAGGGCGCTCAG 10300
CGGGTGTGGCGGTGTGGGGCTGGCTTAATATGCGGCATCAGAGCAGATTGTACTGAGAGTGCAACATATGCGGTGTGAATACCGCACAGATCGCT 10400
AAGGAGAAAATACCGCATCAGGCGGCTTAAGGGCTCTGATACGCTATTTTTATAGTTAATGTATGATAAATGTTTCTTAGAGCTCAGCTGG 10500
CACTTTTGGGGAATGTGCGCGGAACCCCTATTTGTTATTTTCTAAATACATTCAAATATGATCCGCTCATGACACAATAACCTGATAAATGCTT 10600
CAATAATATTGAAAAAGGAGAGTATGAGTATTCAACATTTCCGTGTCGCTTATTTCCCTTTTTTGGGCAATTTTGCCTTCTGTTTGTCTCACCAG 10700
AAACGCTGGTGAAGTAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCTTCGAGAGTTTTCG 10800
CCCCGAAGAACGTTTCCAATGATGAGCACTTTAAAGTTCTGCTATGTGGCGGGTATTATCCCTATTGACCCCGGCAAGAGCAACTCGGTCCGCGC 10900
ATACACTATTCTCAGATGACTTGGTTGAGTACTACCACTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGTGCCATAA 11000
CCATGAGTGATAAAGCTGCGGCCAACTTACTTCTGACAAAGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGATCATGTAACCTG 11100
CCTTGATGTTGGGAACCGGAGCTGAATGAAGCCATACCAACGACGAGCGTGACACCAGATGCTGTAGCAATGGCAACAAGTTGCGCAAACTATTA 11200
ACTGGCGAACTACTTACTTAGCTTCCCGCAACAATTAATAGACTGGATGGAGCGGATAAAGTTGCAAGGACCACTTCTCGCTCGGCCCTTCCGGCTG 11300
GCTGGTATTGCTGATAAATCTGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT 11400
CTACACGACGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCTCACTGATTAAGCATTGGTAACGTGACAGCAAGTT 11500
TACTCATATATCTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGCTGAAGATCTTTTTGATAATCTCATGACCAAAATCCCTTAAC 11600
GTGAGTTTCTGTTCACTGAGCGTCAGACCCGTAAGAAAGATCAAGGATCTTCTTGAGATCTTTTTTCTGCGGTAATCTGCTGTTGCAAAACAA 11700
AAAAACCCGCTACCAAGCGGTGGTTTTGTTTGGGATCAAGAGCTACCAACTCTTTTCCGAAGTAACTGGCTTCAGCAGAGCCGACATACCAATACT 11800
GTCTTCTAGTGTAGCCGAGTTAGGCCACCACTTCAAGACTCTGTAGLACCGCTACATACTCGCTCTGCTAATCTGTTAACCAGTGGCTGCTGCEA 11900
GTGCGGATAAGTCGTGCTTACCGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGTGGGCTGAACGGGGGTTCTGTGCACAGCCCAAG 12000

donderdag, 27 november 1997 10:48
fig 35 pNP8 Map (1 > 12641) Site and Sequence

Page 5

CTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCGGAAGGAGAAAGCGGACAGGTATCCGGTA 12182
ACCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCAGGGGGAAACGCTGGTATCTTTATAGTCTGTGGGTTTCGCCACCTCTGACTTGAGC 12282
GTGGATTTTGTGATGCTCGTCAGGGGGGGGAGCCTATGGAAAAACGCCAGCAACGCGGCTTTTACGGTTCCTGGCCTTTTGTGGCCTTTTGTCTCA 12382
CATGTTCTTCTCTGCTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCCGAGCCGAACGACCGAGCGCAGC 12482
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAACCGCCTCTCCCGCGGCTTGGCCGATTTCATTAATGCAGCTGGCAGCAGGTTTCCCG 12582
ACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTACACTTTATGCTTCCGGCTCGTATGTTGTG 12682
TGAATTGTGAGCGGATAACAATTTACACAGGAACAGCT 12641

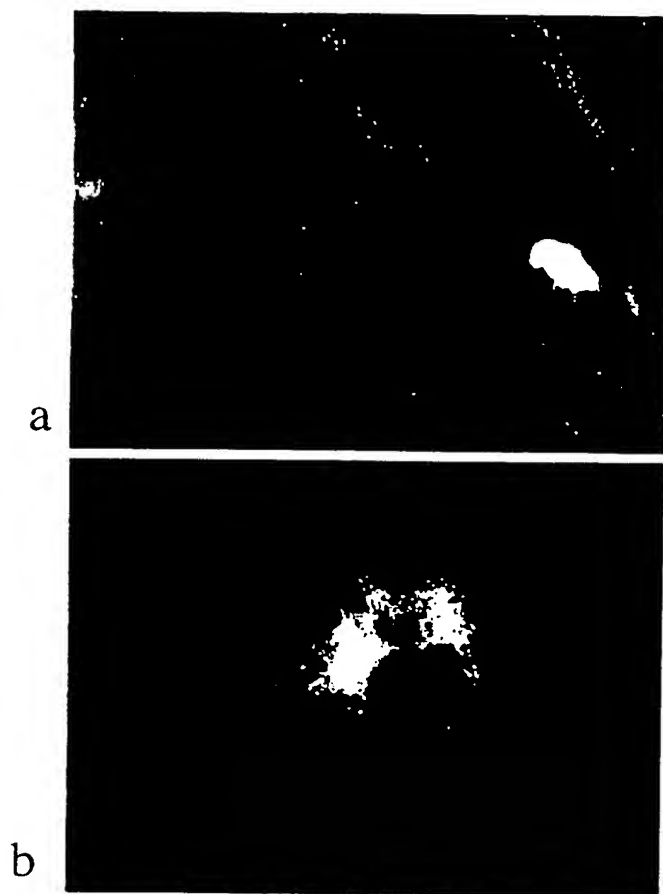


Figure 36: Association of *C. elegans* UNC-53 (expressed from pTB72) with the microtubular cytoskeleton of HepG2 cells. (A) microtubules stained with YL1/2 antibody to tubulin and (B) Staining for *C. elegans* UNC-53.

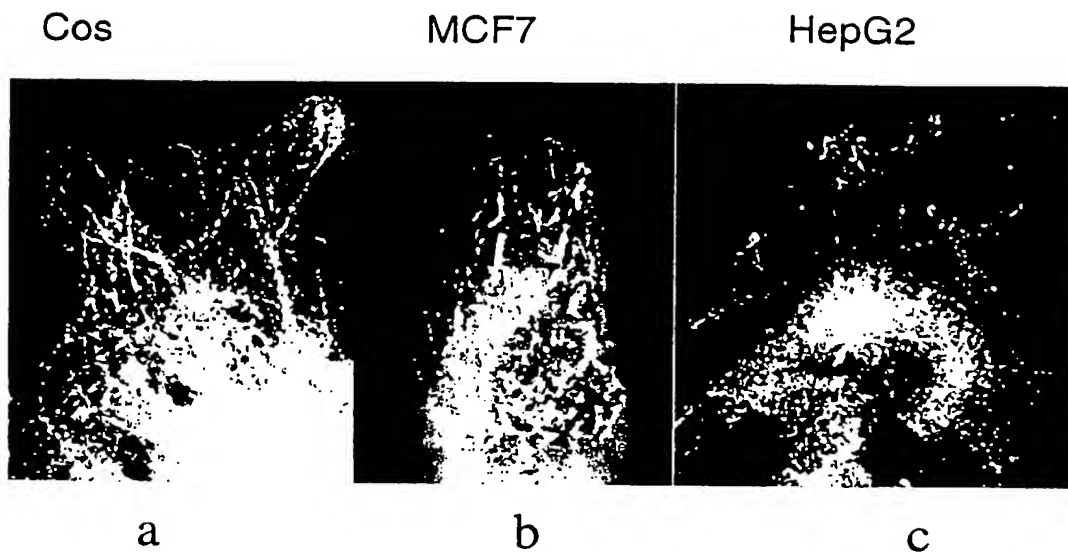


Figure 37: Microtubule (+)-end binding of *C. elegans* UNC-53 following transient transfection with pTB72 of HepG2 (a), MCF7 (b) and Cos cells (c). *C. elegans* UNC-53 was visualised by immunofluorescence using mab16-48.

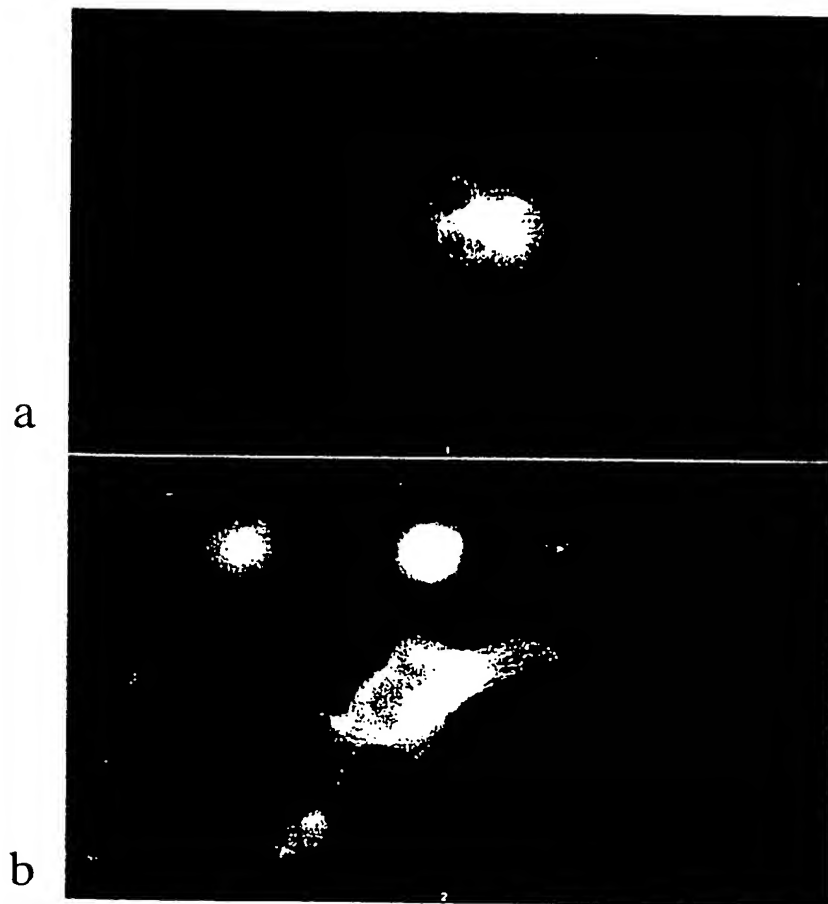


Figure 38

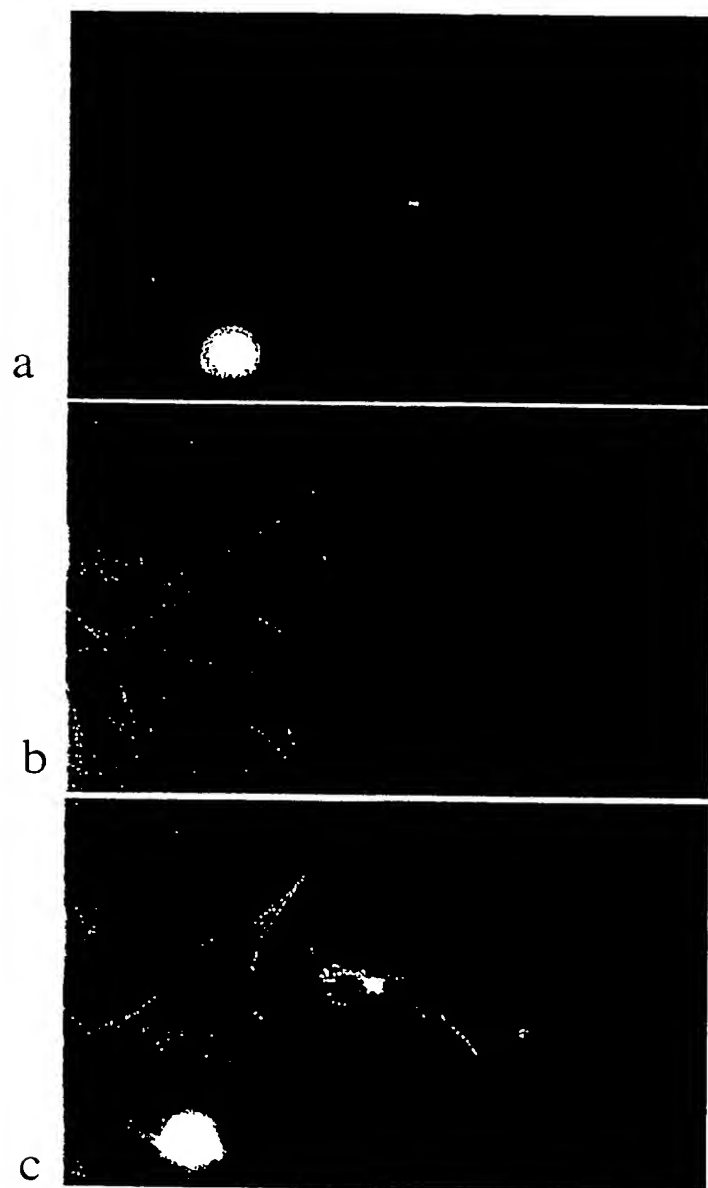


Figure 39

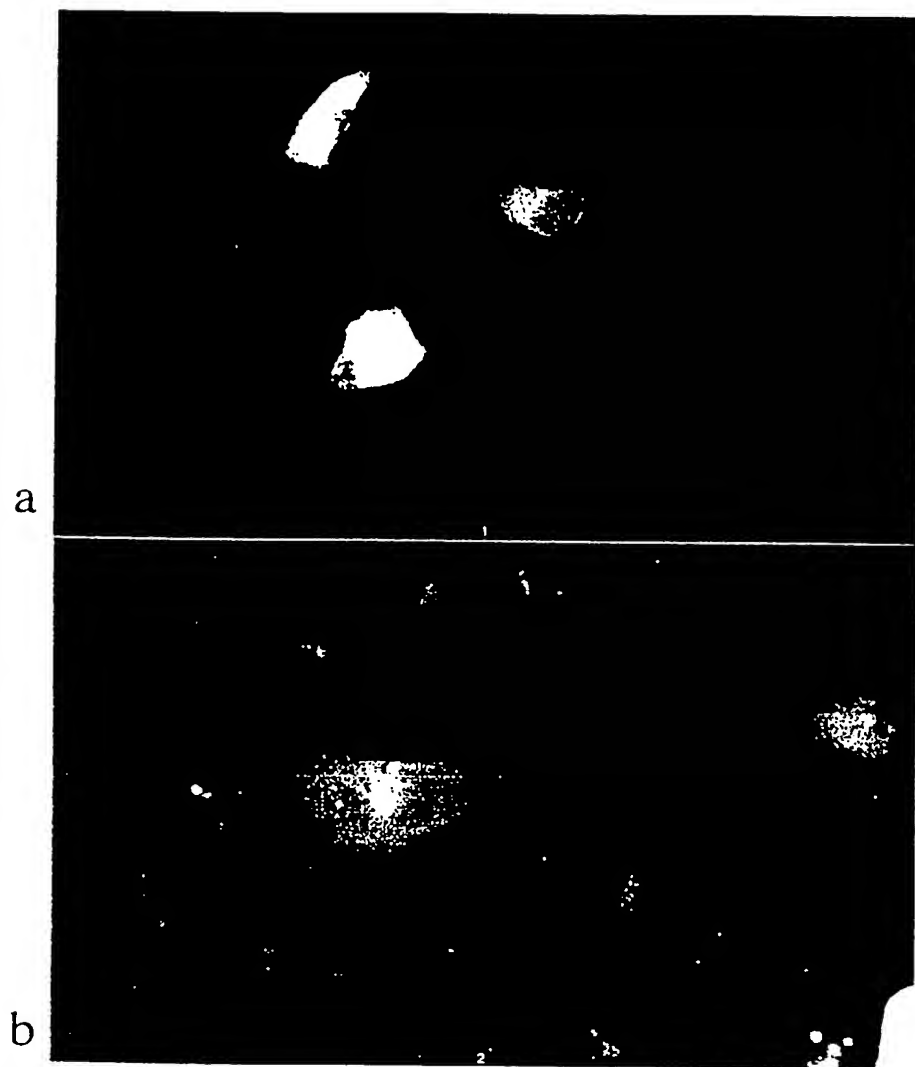


Figure 40

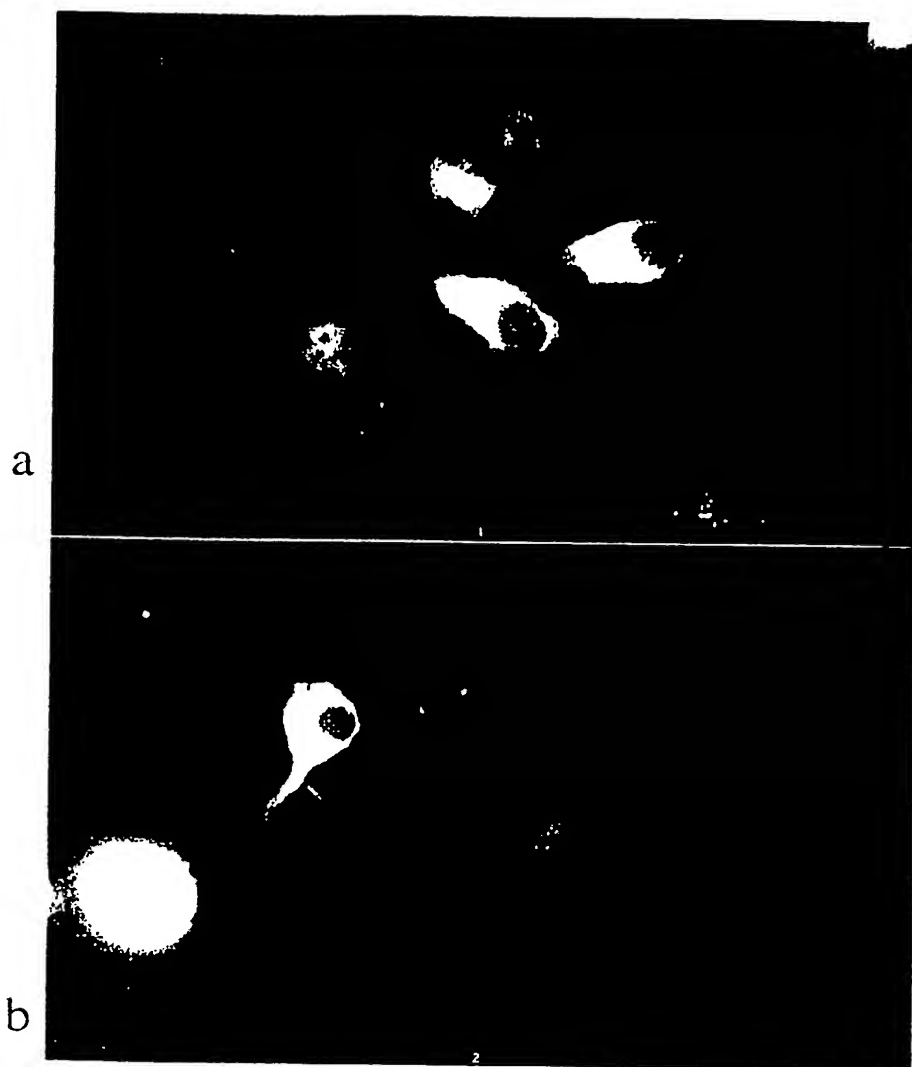


Figure 41

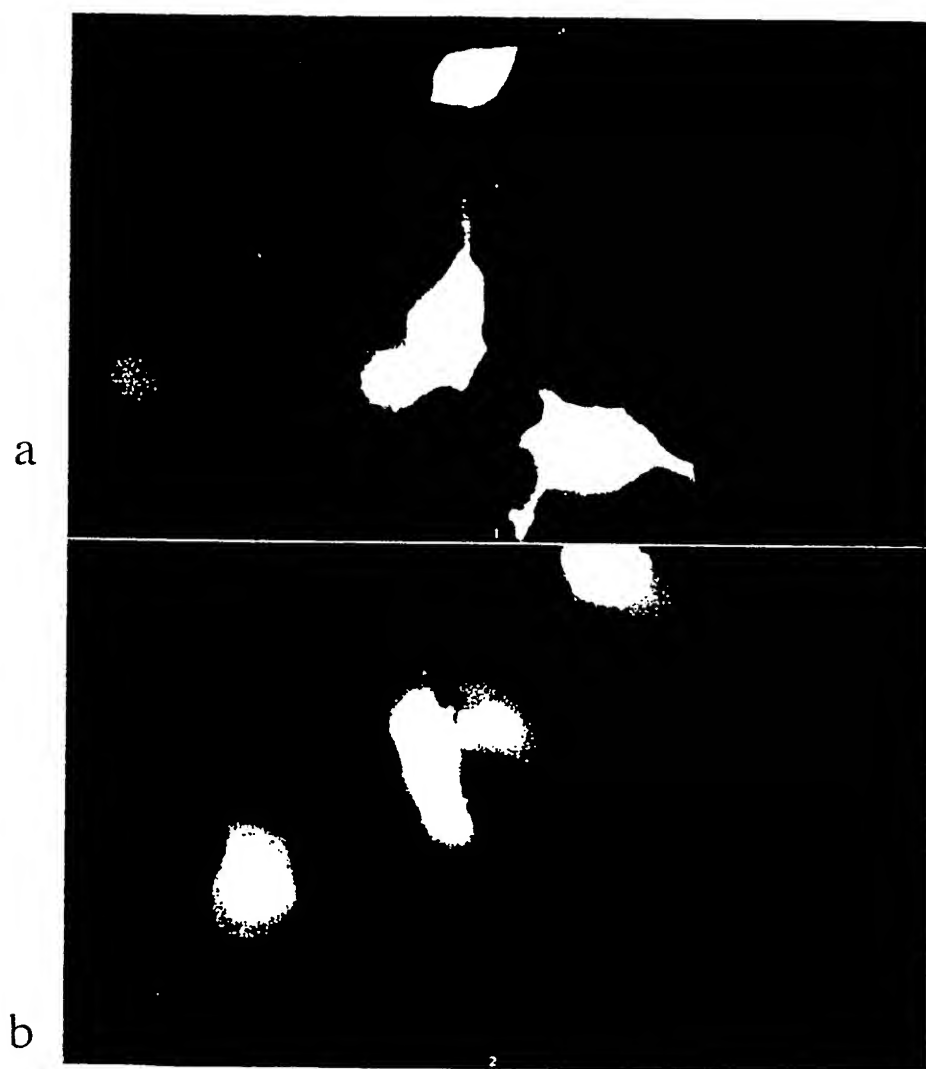


figure 42

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTACGGGAGCCCTTAATGCCTCAGAAACACACCCAAAGAACTTCGGATCAAGAGA
GTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCTC 430

insert pLM1

ORF pLM1

I D F L K K K N S E A Q A V I O G A L N A S E T T P K E L R I K R

CAAACTCTTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAAGAAAAAGAA
GTTTTGAGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTCTTCTTTTTTCT 430

insert pLM1

ORF pLM1

O N S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K

GTGGGTCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCGGATATAGAGGAGATTGCTAC
CAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTCCCGGGTTCAGTCGAAGGAGTATGAGCCTATATCTCTC TAACGATG 440

insert pLM1

ORF pLM1

S V V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T

ACCCGACTCTTCAGCCCCCTCATCCCCAAACTACAGCATGGTTCCACAGAGACTGCTTCACCCCTCCATCAAGTCCTCCACCTTGCTCTCGTGGGCACT
TGGGCTGAGAAGTCGGGGGAGTAGGGGGTTTGATGTCGTACCAAGGTGCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGCCACCCGTGA 450

insert pLM1

ORF pLM1

P D S S A P S S P K L O H G S T E T A S P S I K S S T L S S V G T

GATGTCACCGAGGGCCCTGCTCACCCAGCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGGT
CTACAGTGGCTCCCGGGACGAGTGGGTGCGGGGGTGTGATCCGACAAGGTACGTTTACTCTCTCCTCGGTCTCTTCTCTCCATAGCCTCGACGCGA 460

insert pLM1

ORF pLM1

C V T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R

CTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACAT
GACTCGATACCCCTCTTCCTTACTTCGAATGCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTA 470

insert pLM1

ORF pLM1

S E L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T I H H P

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

GCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCT
CGTCAACCTCCACCTGGACGACTTTCTGTCTTACTGGCTGACTTCCATC3GGGTCCGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGA

480

insert pLM1

ORF pLM1

O L E V D L L K A E N D R L K V A P G P S S G S T P G Q V D G S S

GCATTATCTTCCCCACGCCCTCCTAGGCTGGCACTACCCATTCTTCGGCCCCAGTCTTGACAGACAGACCTGTCACCCATGGATGGCATCAGTA
CGTAATAGAAGGGGTGCGGCAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCCGGGTCAGAACGTCGTGTCTGGACAGTGGGTACCTACCGTAGTCAT

490

insert pLM1

ORF pLM1

A L S S P R R S L G L A L T H S F G P S L A D T D L S P M D G I S

CTTGTGGTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCCGACGACATCATCAAAGGGGACTTGAAGCAGCAGGAATCTTCTCGGG
GAACACCAAGTTTCTCCTTCACTGGGAGGCCACCACTCTACGGGGGCGTGTGTAGTAGTTTCCCTGAAGTTCGTCTCTTAAGAAGGACCC

500

insert pLM1

ORF pLM1

T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q E F F L G

CTGTAGCAAGGTCAAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCAAGTGTCAAGGACTATATTTCTAAATGGACCCAGCCTCTACC
GACATCGTTCCAGTCACCTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCTGTATATAAGATTTTACCTGGGTCGGAGATGG

510

insert pLM1

ORF pLM1

C S K V S G K V D V K M L D E A V F O V F K D Y I S K M D P A S T

CTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAACAGAGTGTGGATGCAGAGCCCCGAGATGCCCTCTTCCGTCGAGGTG
GACCCTGATTCTGTGACTCAGGTAGGTACCGATGTCTGTAGTCGGTGCACCTTCTCACAACCTACGTCCTGGGGGGCTCTACGGAGGAACGGCAGCTCCAC

520

insert pLM1

ORF pLM1

L G L E T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G

TCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACACCTCGGTGTTCGAGACGCTGATCCCAAGCCGATGATGCAGCACTACAT
AGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTACGCAGCTGTCTGGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGTGTAT

530

insert pLM1

ORF pLM1

V N H I S V S L K G L K E K C V D S L V F E T L I P K P M M J H Y I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 9

```

AASCC TCCTGCTGAAGCACCGGCCTCGTCC TCTCGGGCCCCAGCGGCAGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGC
TTCGGAGGACGACTTCGTGGCCGCGAGCAGGAGAGCCCGGGGTCGCCGTGCCGTTCTGGATGGACTGGTTAGCGAACCGGCTCATGEACCACCTGGCG
540
insert pLM1
ORF pLM1
S L L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R
TCTGGCCGTGAGGTCACAGAGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAGA
AGACCGGCAC TCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCTGCAGAACGTTCC TAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCT
550
insert pLM1
ORF pLM1
S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q
TAGACCGGGAACAGGAATGGGGATGTGCCCCCTGGTGATTCATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCAC
ATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACCTCGTCCGAGGTAGTCAC TCAACCAGTTACCCCGGGAGTG
560
insert pLM1
ORF pLM1
I D R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T
CTGCAAGTATCATAAATGTCCCTATATTATAGGTACCAACATCAGCCTGTAAAAATGACACCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACC
GACGTT CATAGTATTTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAACGTGAAC TCGAAGTCTACAAC TGG
570
insert pLM1
ORF pLM1
C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T
TCTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGC
AAGAGGTGTGTGACCTCGGTGCGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTCTCTCG
580
insert pLM1
ORF pLM1
S N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E
TGCTTCGGGTGCTCGACTGGGTACCCAAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACC TCAGACTTCCTCATCGGCCCTTGCTTCTT
ACGAAGCCCAAGAGCTGACCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTCTCGTGTCTGGAGTCTGAAGGAGTAGCCGGGAACGAAGA
590
insert pLM1
ORF pLM1
L L R V L D V V P K L V Y H L H T F L E K H S T S D F L I G P C F F

```

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 10

TCTGTCGTGTCCTTGGCATTGAGGACTTCCGGACCTGGTTTCATTGACC TGTGGAACAACCTATCATTCCTTATCTACAGGAAGGAGCCAAAGATGGG
AGACAGCACAGGGTAACCGTAAC TCC TGAAGGCC TGGACCAAGTAAC TGGACACCTTGTGAGATAGTAAGGGATAGATGTCTTCC TCGGTTCTCTAACC
+ 810
----- insert pLM1 -----
----- ORF pLM1 -----
L S C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G
ATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTG
TATTTCCAGGTACCTGTCTTTCGACGAACCCTCCTGGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTGTCTTGGTGTAGTTTCGACA
+ 810
----- insert pLM1 -----
----- ORF pLM1 -----
I K V H G Q K A A V E D P V E V V R D T L P V P S A Q Q D Q S K L
ACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTC TCGACTCAGATCC
TGGTGGACGGGGTGGGTGGCACCCGGGAGTGTGTAACGGAGTGGAGGGCTCCTATCTGTCTAGTTTCTGTCTGGGGTTCAAGAGACCTGAGTCTAGG
+ 820
----- insert pLM1 -----
----- ORF pLM1 -----
Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S D P
TCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCTCGACCCCAACCTTCAGGCAACACTTTAA
AGACTACCGGTACGACGACTTTGAAGTCTTCGACGGTTGATGTAAC TCAAGGCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAAATT
+ 830
----- insert pLM1 -----
----- ORF pLM1 -----
L M A M L L K L Q E A A N Y I E S P D R E T I L D P N L Q A T L
GGGTTCCGCAATCACTGTCAACCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCCTCTCTCCCTCTCTCTTTTCAGAGCACTGGCTCTCCAG
CCCAAGCCGTTAGTGACAGTGGGGGCTGTCTGTCTTGCACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTC
+ 840
----- insert pLM1 -----
G F G N H C H P R T A E R V H O L S . L L L S P L L F O S T G S P
CCCCAGGAGGAGAACAGGAGGGAGGAGAGATGAAAGAGGAGGGACAGGTCTTGGTGCTGTACCTTTGAGAACTTCTTAGGAAGGAATGGTGGGGTGGC
GGGTCTCTCTTTGTCCTCCTCTCTCTTACTTTCTCTCTCTGTCCAAGAACCACGACATGGAAACTCTTGAAGGATCTTCTTACCACCCACCG
+ 850
----- insert pLM1 -----
A P G G E Q E G G G D E R G G T G S V C C T F E N F L G R N G G V A
GTTTGGGAAC TTGTGCCCCCTAAACACATTTACTGGCTCTCTAATGACTTTGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTCATTT
CAAACCTTGAACACGGGGGATTGTGTAATGACCGGAGGAGATTACTGAAACCCCTTTTCTACTAAGACCCAGAAAGGGAAC TGAAGAACAAGTTAA
+ 860
----- insert pLM1 -----
F G N L C P L N T F T G L L . L V G K D D S G S F P . L L V S I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 11

ACAACTCTCGGCTTTCTGGGGAGGGGTTGAGAAACATCAAAACACTGCAGCAGTTCCCGGAATTCAGCTTGGACTTAACAGGCTGAACCTTGCTCA
TGTGAGGACCCGAAAGACCCCTCCCAAGTCTTTGTAGTTTGTGACGTCGTCAGGGGCCCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGT 570

insert pLM1

T N S V A F V G G V Q K T S K H C S S S P E F S L D L T R L N L L

AAAGAAGCCGAATTCAGCACACTGGCGGCCGTTACTAGTTCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTACTTGCTTTAAAA
TTTCTTCGGCTTAAGGTCGTGACCGCCGCAATGATCAAGATCTATTGACTAGTATTAGTCGGTATGGGTAAACATCTCCAAATGAACGAAATTTT 580

insert pLM1

K R S R I P A H W R P L L V L D N . S . S A I P H L . R F Y L L .

AACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAAATAAGCAATA
TTGGAGGGTGTGGAGGGGACCTGGACTTTGTATTTACTTACGTTAACAACAACAATGAACAAATAACGTCGAATATTACCAATGTTTATTTTCGTTAT 590

T S H T S P . T . N I K . M Q L L L L T C L L Q L I M V T N K A I

GCATCACAATTTTCAAAATAAGCATTTTTTTCACTGCATTCTAGTTGTTGGTTTGTCCAACTCATCAATGTATCTTAACGCGTAAATGTAAGCGTTA
CGTAGTGTTTAAAGTGTATTATTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATTGCGCATTAAACATTCGCAAT 700

II ori

A S Q I S Q I K H F F H C I L V V V C P N S S M Y L N A . I V S V

ATATTTTGTTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGAC
TATAAAACAATTTAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCCGTTTATAGGAATATTAGTTTCTTATCTG 710

II ori

N I L L K F A L N F C . I S S F F N Q . A E I G K I P Y K S K E . T

CGAGATAGGGTTGAGTGTGTTCAGTTTGGAAACAAGAGTCCACTATTAAAGAAGCTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGAT
GCTCTATCCCAACTCACAACAAGGTCAAACCTTGTCTCAGGTGATAATTCTTGCACCTGAGGTTGCAGTTTCCCGCTTTTGGCAGATAGTCCCGCTA 720

II ori

E I G L S V V P V W N K S P L L K N V D S N V K G R K T V Y Q G D

GGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGAGCCCCGATTAGAGCTT
CCGGTGATGCACTTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGTGATTAGCTTGGGATTTCCTTCGGGGCTAAATCTCGAA 730

II ori

G P L R E P S P . S S F L G S R C R K A L N R N P K G S P R F R A

GACGGGAAAGCGCGCAACGTGGCGAGAAAGGAAGGAAGCAAGGAGCGGGCGCTAGGGCGCTGGCAAGGTAGCGGTACGCTGCGCGTAAC
CTGCCCCTTTCGGCCGCTTGACCGCTCTTTCTTCTCTTCTTCTGCTTTCCTCGCCCGGATCCCGGACGTTACATCGCCAGTGCGACGCGCATG 740

II ori

R G K P A N V A R K E G K K A K G A G A R A L A S V A V T L R V T

CACCACCCCGCGGCTTAATGCGCCGCTACAGGGCGGCTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTTATTTTCTAAATA
GTGGTGTGGGCGGCGCAATTACGCGCGATGTCCCGCGATCCACGTGAAAAGCCCTTTACACGCGCTTGGGGATAAACAATAAAAGATTAT 750

II ori

T T P A A L N A P L O G A S G G T F R G N V R G T P I C L F F . I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 12

CATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTGAGGCGGAAAGAACCAGCTGTGGAA
GTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTACGAAGTATTATAACTTTTCTCTTCAGGACTCCGCCCTTTCTTGGTCGACACCTT
H S N M Y P L M R Q . P . . M L Q . Y . K R K S P E A E R T S C G 760
TGTTGTCTAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCC
ACACACAGTCAATCCACACCTTTTCAGGGGTCCGAGGGGTCTGCTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTTCAGS 770
M C V S . G V E S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P
CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTAACTCCGC 780
GGTCCGAGGGGTCTGCTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCG
Q A P Q Q A E V C K A C I S I S Q Q P . S R P . L R P S R P . L R . 790
CCAGTTCGCCCATTCGCCCCCATGGCTGACTAATTTTTTTTATTTATGCAAGGGCCGAGGCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGG
GGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAAAAAAATAAATACGCTCCGGCTCCGGCGGAGCCGGAGACTCGATAAGGTCTTCATCACTCC 790
P V P P I L R P M A D . F F L F M Q R P R P R P L S Y S R S S E
AGGCTTTTTTGGAGGCCTAGGCTTTTCAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC 800
TCCGAAAAAACCTCCGGATCCGAAACGTTTCTAGCTAGTTCTCTGTCTACTCTAGCAAAGCGTACTAACTTGTTCACCTAACGTGCGTCCAAGAGG
E A F L E A . A F A K I D Q E T G . G S F R M I E Q D G L H A G S P Kan/Neo
GGCCGCTTGGGTGGAGAGGCTATTCCGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCCGCGTGTTCGGCTGTCAAGCGAGGGGCGCCG 810
CCGGCGAACCCACCTCTCCGATAAGCCGATACTGACCCGTTGTCTGTTAGCCGACGAGACTACGGCGGCACAAGGCCGACAGTCCGCTCCCGCGGGC
A A V V E R L F G Y D V A Q Q T I G C S D A A V F R L S A Q G R P Kan/Neo
GTTCTTTTTTGAAGACCGACCTGTCCGGTGGCTGAATGAAGTGAAGACGAGGCGCGGCTATCGTGGCTGGCCACGACGGGCTTCTTGGCGAG 820
CAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGGCTC
V L F V K T D L S G A L N E L Q D E A A R L S V L A T T G V P C A Kan/Neo
CTGTGCTCGACGTTGTCTGAAGCGGGAAGGACTGGCTGCTATTGGGCGAAGTGCCGGGCGAGGATCTCCTGTCTCTCACCTTGCTCTGCCGAGAA 830
GACACGAGCTGCAACAGTGACTTCGCCCCTCCCTGACCGACGATAACCCGCTTCACGCGCCCGTCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTT
A V L D V V T E A G R D V L L L G E V P G Q D L L S S H L A P A E Kan/Neo
AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGT 840
TCATAGGTAGTACCGACTACGTTACGCCGCGACGATGCGAATAGGCCGATGGACGGTAAGCTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCA
V S I M A D A M R R L H T L D P A T C P F D H Q A K H R I E R A R Kan/Neo
ACTCGGATGGAAGCCGGTCTTGTCTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTTCGCCAGGC TCAAGGCGAGCATGC 850
TGAGCCTACCTTCGGCCAGAACAGCTAGTCTCTAGACCTGCTCTCGTAGTCCCGAGCGGGTGGCTTGACAAGCGGTCCGAGTTCCGCTCGTACG
T R M E A G L V D Q D D L D E E H Q G L A P A E L F A R L K A S H Kan/Neo

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 13

CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTATCGACTGTGGCCGGC- 860
GGCTGCCGCTCCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGCGAAAGACCTAAGTAGCTGACACCGGCCGA
Kan/Neo
P D G E D L V V T H G D A C L P N I M V E N G R F S G F I D C G R L
GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGTGCTTTACGG- 870
CCCACACCGCTGGCGATAGTCTGTATCGCAACCGATGGGCACATAACGACTTCTCGAACCGCGCTTACCGGACTGGCGAAGGAGCACGAAATGCCA
Kan/Neo
G V A D R Y Q D I A L A T R D I A E E L G G E W A D R F L V L Y G
ATCGCCGCTCCCGATTGCGAGCGCATCGCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC 880
TAGCGGCGAGGGCTAAGCGTCCGTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCTGCTGGC
Kan/Neo
I A A P D S Q R I A F Y R L L D E F F . A G L V G S K . P T K R R
CAACCTGCCATCAGGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACCGCGCTGGATGATCTCCAGCGC 890
GTTGGACGGTAGTGCTCAAAGCTAAGGTGGCGGCGGAAGATACTTCCAACCGAAGCCTTAGCAAAAGGCCCTGCGGCGGACCTACTAGGAGGTCCGC
P T C H H E I S I P P P S M K G V A S E S F S G T P A G . S S S A
GGGGATCTCATGCTGGAGTTCTTCGCCACCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAA 900
CCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCTTCTTGGGCGCGATATGCCGTTATT
G I S C V S S S P T L G G G . L K H G R R Q Y R K E P A L . R Q .
AAAGACAGAATAAACGACGGTGTGGGTCGTTTGTTCATAAACCGGGGTTGGTCCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCCATTC 910
TTTCTGTCTTATTTTGGGTGCCACACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCAGCGTGAGACAGCTATGGGGTGGCTCTGGGGTAAC
K D R I K R T V L G R L F I N A G F G P R A G T L S I P H R D P I
GGGCCAATACGCCCGGTTCTTCTTTTCCCCACCCACCCCAAGTTCCGGTGAAGGCCAGGGCTCGCAGCCAACGTCGGGCGGCGAGGCCCTGCC 920
CCCGGTTATCGGGCGCAAGAAGGAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCGGAGCGTGGTTGCAGCCCCCGCGTCCGGGACGG
G A N T P A F L P F P P P T P Q V R V K A Q G S Q P T S G R Q A L P
ATAGCCTCAGGTACTCATATATACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCA 930
TATCGGAGTCCAATGAGTATATATGAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACTTCTAGGAAAAATATTAGAGTACTGGT
P Q V T H I Y F R L I . N F I F N L K G S R . R S F L I I S . P
AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGTAAGAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTG 940
TTTAGGGAATTGCACTCAAAAGCAAGGTGACTCAGCTGCGGCATCTTTCTAGTTTCTAGAAGAACTCTAGGAAAAAAGACGCGCATTAGACGAC
pUC ori
K S L N V S F R S T E R O T P . K R S K D L L E I L F F C A . S A
CTTGCAACAAAAAACCCGCTACAGCGGTGGTTTGTGGCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCGCAE 950
GAACGTTTGTTTTTTGGTGGCGATGGTCCGCCACCAACAAACGGCCTAGTTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGCTCGCGTC
pUC ori
A C K Q K N H R Y Q R V F V C R I K S Y Q L F F R R . L A S A E R R

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 14

ATACCAAACTACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCTACATACCTCGCTCTGCTAATCTGTACCAG
TATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGT
950
pUC ori
Y Q I L S F . C S R S . A T T S R T L . H R L H T S L C . S C Y Q
TGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGCTGAACGGGGGTTTCGTG
ACCGACGACGGTCACCGCTATTACGACAGAAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCAC
970
pUC ori
V L L P V A I S R V L P G V T O D D S Y R I R R S G R A E R G V R
CACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGCGGGAC
GTGTGTCGGGTGCAACCTCGCTTGTGGATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTCCGCTG
980
pUC ori
A H S P A V S E R P T P N . D T Y S V S Y E K A P R F P K G E R R T
AGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCCTGGTATCTTTATAGTCCGTGTCGGGTTTCGCCACC
TCCATAGGCCATTGCGCGTCCCGAGCTTGTCTCTCGCGTGTCTCTCGAAGGTCCCCCTTTCGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGG
990
pUC ori
G I R . A A G S E Q E S A R G S F Q G E T P G I F I V L S G F A T
TCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGGGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTACGGTTCTTGCCCTTTTGCTG
AGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGTCTGTGCGCGGAAAAATGCCAAGGACCGGAAAAACGAC
1000
pUC ori
S D L S V D F C D A R Q G G G A Y G K T P A T R P F Y G S V P F A
GCCTTTTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CGGAAAACGAGTGTACAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA 10070
G L L L T C S F L R Y P L I L V I T V L P P C I

donderdag, 27 november 1997 16:48
fig 35 pNP8 Map (1 > 12641) Size and Sequence
Enzymes: All 146 enzymes (No Filter)
Settings: Circular, Certain Sites Only, Standard Genetic Code

Page 1

ATGACCATGATTACGCCAAGCTTGCAATGCTGCAGGAATTCGATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAAGAAATTGGAGCAACTACC 100
CACATCCATTATGCCACCCGCGGTTCTAAGTGAGTTAATTTTGAGTTTACGACTACAAAATGTGTTCTTTAATAACTATCTTCGACTTGAGTCTATT 200
CTGTATGACTAGTTGTTGAGTGATTTTTCATTGAGAAAATATTAAGGAACATTATTTACTTGTCTATTGCCCCTAAGTTGATTTAGTTTTCGATC 300
AACTAGATCTTCAAAAATTGCAATACAATTCATTTTCAGATTACCTCGCCACGTGTCGCCACGTGAGCAACCGCTTCAGCAACTAACCCAAATTC 400
ACTTTCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAAATATCGAAAATTTGTAAGAATTTTATTTTGAGCTCAAACTGTATAAAATGCC 500
CAGAAAAGAAGATGATAAAATGTAGTTTTTTTGCAAAATTCACCTTTATGCTCTAATATGACGGCTTATATCTCAATTTTCTTGAGTTTTATCAA 600
AAATTTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTGTGAGTTGACAGCTTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTAA 700
AGTGAAGGAGCGCAAGTCTATCTGGAATAATGATCTGAACAAATTTGTGCTATTCTCAAAATGTTAAGACATGTTTTGAAGATTTTTTCAAATTCG 800
CACTAGTTTCAGAACCTTCTCTTTTGTATGAAAAGTAAAAAAAATCTTTTCAAACTCCACCGCCACCATGTTTCAACTCTTAATTTTATAAAATTT 900
TGCAATTTACAAATCGCTCCCTCTTGCCTGAAAAGTGCCTCAAAATCAATTTCTCGGCTTCATAATGACTTTTAAATGATGTGAGAAAACACAGAAG 1000
AGGCTAACTAAATGACAGGGACAGGTTGTCT 1100
TTGTCCATTTTGCTTATAAATTTGTGTGTGGAAGGAACTACACGGGAGACGGTCAATTAATTCGAATGAGAGCATGGCAATTAATCTTTTGGAAAT 1200
TGATGAATAAGATAGAGCGGATGACACTGGCTGGTAGTAGTATGAGTGTAGAATTGTTTTTATCGTCTCAACTTGGCATGAGTCTTCCCCGCTCT 1300
CATCACTGACAATTAATGTCTGGGTTTTATCGGCTCTTCTCTATTCGGCACTATTCTGGGTACCAAACTGGAATACATTTTACTACTATTCAAGCC 1400
ATTTATTTTGATATTTAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAGTTTATTTAAAAAACCATATTTTCGATTTTAAAAATCTGAAAAGT 1500
TTCAAAAATCAATAAATATTCCTAACAAATGTATGGCTAAATTTTATTTCTACTGTTGACAAATCTTTATATGTATCACTGTTTCCATCTCAAA 1600
ACCTTGAAATCCCCAAGTTATAGGAAGCTCCGTGTCAATTTCCATGCTATGAATGCTACTCAGCACATATCAAAAATTAAGCTAGAGCGTTGATAA 1700
TTATTTGGGACCGTAATAAAGTGCAAGCAGTTAGAATTTTAATTCAGCACAGATTATCTCAAAATCAATCTTTGAACATTGAGCAGTTCTGACAA 1800
TTTTCCATGCTTTTGGCCATTAAAAACTTTCTCACTCTTCATCCATCTCACTCGTATCAAAAAAGTATAGCAAAAGCCGACTCTACTTTTAAAG 1900
AGAAGGAGATCTAGGACACATGGCGTGTGACCTTTTCTATCTGTCGGTTCGCTCAAAATCACGCTCATACTAACTCTTCAAAATAGCCATAGACCTC 2000
CTGTTTTCTTCTCGTTTGACTCGCGCTATTTTGTGCTGCTGAAAGCCGGGAAAATTTAGTATTTATGAGCTTATCTTTATGCAATACATA 2100
AAAAACGAGGCAATTTAAAAATATTAATAAATTAATGAGGTTGTAGATGTAGATTTGCAAAAGAGAAAAAACAACAAATAGCAACCGCCAGATCAAAA 2200
TTCTATTTAAAGGTTTTCAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAATCTGAAGTGCCTGCAATAATCTAGTGTAACTTTAGATTGAACTCGGAA 2300
ATCTAAGCCTGAACATATAGCTTATTTCTAGATCTTAGTTGCGCATAGCTCAAGCCCAAGCAGAAATGACTTGCAATTTAGTTAAGCCTAGATTGACTT 2400
GCTTGCTTCAGTCTAATCCAGACTAGATTTCCAGAGAGTTTCAATTTTAAATGTTTCCAGTTTCTGTTACTTAAATCTTAATGCCCTGTGATGCGT 2500
AAAATCGTTATCCCTTTCTCTCACACTTTCAATTACAGATTCATCAAGATTGCTATCAAGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAAC 2600
ACTTCATCAAAATATACAAATTCATTCCGTCCGTGAGCGGTTGAGTGGCAATAAATGTTGCTGACCATATCCACATCTGCGAAGAGCTTAGGTA 2700
TCCGATCTTCCGGCTTCTTTTAAAGATTAATTTTCAAGATCATCAACGTAAGCTCTATTTTGAATCTAAACCGACTACCTCCCACTCCA 2800
AAAACTTCTAGACCAAAACCCAGCTAGTTGCTGTGCTACAACTACAAAATCGGAAGCTCAAGCTAGCCGCTCCGAAGCGGTGAGACCCCAAAA 2900
CTGCTTCTGTGAAGACTATTGGAGCAAAACAAGACCCGATAACAGCGGTGCTGCTGCTGCTGGAATGCTGAAATTAAGTTATTCAGTAGCAAAAACC 3000

donderdag, 27 november 1997 18:48
In 35 pNPB M30 (1 > 12641) Site and Sequence

Page 2

CATCTTCCTCATCGAATAGCCCAACCTACGAGAAAGCGCGCGCTCAACAACAACTTTGTGAAATCGCTGCCCAAGTGAAGTGGCCT 3100
GAAGCGCGGACAGTAAGCTGGGAAGTGCCACGTCTATGTGGAAGCTTTGTACGGTGAGTATTTGAAATCGGAAATGGGAATGTATTTTAAAAACT 3200
GAAATTTCTACAAAATAAATAAATAAAGATTTTTCTCTGATAGTATTGCATCCACTATTTTACTTTGAAGATTTATATCTTGTTTCATATTGAAG 3300
ATATCAGATATAGAAAAGAAATAAATAATTTTGACAGTTGATAATTTTGTATAGGACCAAGACAAGTGAGATATAAGCTGTCAAAGTTGATTTTC 3400
AAGAAATTTTAAACCTAGTTTTCGGAAGCTCTGGGCTCATCTATTTAGAACCGATTCGTAATCTTCCGTTCTTGACTCTACCAAAACCAAAA 3500
CCAACCTACTAATAAAATGATGAGACAATGGGAATTGTCTCCATTTTCTCTTCTCTTGACACTCTTCAGAACTATGTCCCATCTTTTGTGCT 3600
TTGTGTCCTCCCATAAAGACTCTTCCGGAATAATGTTGCAACGGAAGTGATAATTCGAGCATTTTTCGACGTGAGGGCCGAAAAACACATCTGGCTGA 3700
CAAGAGTAAAGCAATTTCTCAGCTCTTCTTCCGCTTTTCAATCTGTTTTCAAAATGAGCTACTACAGAGTGAAAGAGCACAAATTGCAAAACATT 3800
TTTGTGTGAGATGCACTTTTGAATAATTAATCTTACGTTTTAGTTTCTAGTATTTATTTTTCATATAAATTAGAGCTTCTTAGACCTCTATATTT 3900
TTTAAACTTCTACTGAAATATACGAGATTCTTTGACTTTCGGAATTGTCTTATGGCTTCTATTATTTATGAGAAAACATTTTAAAAATTTTTT 4000
GAAAAAAACTGTGCATCTTCTTTTTTACATAGTAATTTCCAGCCAAAAGTTTCTACCGTAAACGGACGCCCAATCATATCTCAACAGACTCSA 4100
AACGATGCTCAAGAGCAGTGAAGAAGAGTCCGGATACGCTGGATTCAACAGCAGCTGCCAACGTCAATCATCGACGGAAGGTTCCCTAAGCATGCATTC 4200
CACATCTTCCAAGGTTCTGTTTATAGGAGAACTGTTTGTGTTTCTGACCTTACATAGTCTCGGATGTTTATAAAGTGAGGTCTCTGGGACAC 4300
CTGCCATAAATGTGAATCCGCCATTTGTTGTTACAAAACCTTTGACAGCACCTGCTTATACATTTTATGGATAAATGTACATCGGTATTTGTCAA 4400
ACCCAACTTTTAAATTTTATTTTACATCAAAAATGATGTTAAAGTTTAAAGATATTTACGAAAAATGTTTACTTAAACTTTTATATCGATA 4500
AAATTTTGAACATTCAGATAGGAGTTCCGTCCCTAAAACCTTTGTGTCGCGGAGAGCTTTTGAATTAATCTCATTTTATAAGTTCGAACTAAT 4600
TTTTTGTGCAACTGAATTTTACAGATGAACTTTAAGTTTCAATTTATCCCATTTGAAACCGTCCCTTCTATAAACTTCAAAATTTTACAGATTCAA 4700
CGTCAGACGAAAAGTCTCCGTATCAGACGATCTTACTCTTAACGCTCCATCGTGACAGCTATCAGACAGCGATAGCCGCAACACCGGTTTCTCCAAA 4800
TATTATCAACAAGCTGTTGAGGTGAGTATTTTGTGTTCTGGGTAGAGGCTTCTGTCAAGTTTGGCTAAATTAATTAATCTTGTGTTAGAGGCTGGC 4900
AAAGCCATTGATCAAGCATGGGCTAACTGGGCGGCTGAACCATGTACATATCTTTGGCCGAGTAGTTGCAATCTAAAGATTGGAAGCTGGCTTCA 5000
AAGTCGGACTAGGCAAAAGTGCAAAAATGGAATAATCTTGAATTCATACGCTTCCGTCTTCCATCTTCTTTTTTGTCTCTTTTGTGCGAGAT 5100
TTTCCCTTTTATGATTTTAAAGTTTGTATACCTTAATGTCTGGCTTCCCTTCTAAGAGCTTCTATATTTTGAATAAATCAATTTTATAGGA 5200
AAAACCAACTGGCAGTGAAAGGAGTGAAAGGACAGCGAAAAAGATCCACTCCAGCTGTTCGGCCAGTGACACCCAGCCAAATCGGAGTTGTT 5300
AGTCCAATTATGGCACAATAAGAGTTGACAAATGGTACGTTTATCTGAACTTTACTTATGTTTGGTGGTGACGTTTTTGTGACCATGTGATGGGA 5400
AGTAATTTTGGATTATTTAAAGTTTGTGCGGGAATAGTAAGGAGGAGTACAATTTATTTTGTAGAGGTTTCAAGTAACTTTGATTTTCTGACCA 5500
TAAGTTTCTTCTGAAAGTTGTTTAAAAAATTGAGTTAAAAATAAATAAATCTCTATAAAAAATTTCTAAATCTTGGGAATTTTCTCAAAAATGTT 5600
TTTCCAAATATGTCTAATAGTAAGATTTGTTGTGAATTAACAACATATTTTAAATAATTTTAAATTTATTCGATTTTCTGTTCCGAGGAAACGACA 5700
AAAAATCAGAAAAAGCGAAATTAATTCAAAAAATAATTTTGAAACTCAAAAATAAGCTACTTTCAAAAAATCAACAAAAAATATCAAAAGAT 5800
TCATATTTTCAGAAATAGAGACATAGTCAAAAATATCAAAAAATCACTATTTTCCGGAATAAATGAGAAAAATCCAAAAATGTAAAAACAAAA 5900
TTGAGAAAAATCCAGAAATGAAAAAATTCCTTGAGGAAAAATTAATAATTTTAAATGTGTGATTTCTGAAACCAAGCATTTCCGACTTTCCGGC 6000

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 3

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTTCTAAATTACCCTCGCCACGTGTCGCCACGTGAGCAACCGCTTCA
TTTCTAGTCTTCTTTAACCTCGTTGATGGGTGTAGGTAATACGGTGGGCGCCAAAGATTTAATGGGAGCGGTGCACAGCGGTGCAGTCGTTCGCGAAG 200

eGFPC.e.unc53xba

C.e.unc53 xba

K D Q K K L E Q L P T S I M P P A V S K L P S P R V A T S A T A S

GCAACTAACCCAAATTCCAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAAATTGATTCATCAAAGATTGGTATCA 1800

CGTTGATTGGGTTTAAGGTTGAAAGGTGTTTACAGTTGTAGGTCCGAAGTCTGAGGTGTGAGTCTTATAGCTTTTAACTAAGTAGTTTCTAACCATAGT

eGFPC.e.unc53xba

C.e.unc53 xba

A T N P N S N F P Q M S T S R L Q T P Q S R I S K I D S S K I G I

AGCCAAAGACGTCTGGACTTAAACCACCCTCATCATCAACCACTTCATCAAATAATACAAATTCATTCCGTCGCGAGCCGTTTCGAGTGGCAATAATAA 2000

TCGGTTTCTGCAGACCTGAATTTGGTGGGAGTAGTAGTTGGTGAAGTAGTTTATTATGTTTAAAGTAAGGCAGGCAGCTCGGCAAGCTCACCGTTATTATT

eGFPC.e.unc53xba

C.e.unc53 xba

K P K T S G L K P P S S S T T S S N N T N S F R P S S R S S G N N N

TGTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGAATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCAAAAA 2100

ACAACCGAGCTGCTATAGGTGTAGAGCTTCTCGAATCTTAGTAGTAGTTGCATGTGAGATAAAGCTTAGATTGGCTGGATGGAGGGTTGAGGTTTTT

eGFPC.e.unc53xba

C.e.unc53 xba

V G S T I S T S A K S L E S S S T Y S S I S N L N R P T S Q L Q I

CCTTGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTTCCTTTAAAAAACCTCCACACCTCCCTCTGAA 2200

GGAAACCTTAGGTGGCTAGATCTATTGACTAGTATTAGTCGGTATGGTGTAACATCTCCAAAAAGAAGAAATTTTTTGGAGGGTGTGGAGGGGAGCT

eGFPC.e.unc53xba

V D P P D L D N . S . S A I P H L . R F Y L L . K T S H T S P .

CCTGAACATAAAATGAATSCAATTGTTGTTGTTAACTTGTATTGTCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAA 2300

GGACTTTGTATTTTACTTACGTTAACAACAACAATTGAACAAATAACGTCGAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTATTAT

T . N I K . M O L L L L T C L L O L I M V T N K A I A S O I S O I I

GCATTTTTTCTACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAACGCGTAAATTTGTAAGCGTTAATTTTTGTTAAATTCGCGTTA 2400

CGTAAAAAAGTGACGTAAAGTCAACACCAACAGGTTTGAGTAGTTACATAGAATTGCGCATTTAATTCGCAATTATAAAACAATTTAAGCGCAE

M F F H C I L V V V C P N S S M Y L N A . I V S V N I L L K F A L

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 4

AATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTGTTTC
TTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCCGTTTGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCAACTCACAACAAAS
N F C . I S S F F N Q . A E I G K I P Y K S K E . T E I G L S V V 150

CAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACC
GTCAAACCTTGTTCAGGTGATAATTTCTTGACCTGAGGTTGAGTTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACTTGGTAGTGG
P V V N K S P L L K N V D S N V K G R K T V Y Q G D G P L R E P S P 250

CTAATCAAGTTTTTTGGGTCGAGGTGCCGTAAGCACTAAATCGGAACCCCTAAAGGGAGCCCCGATTAGAGCTTGACGGGAAAGCCGGCAACGTG
GATTAGTTCAAAAAACCCAGCTCCACGGCATTTCGTGATTAGCTTGGGATTCCCTCGGGGGCTAAATCTCGAACTGCCCTTTCGGCCGCTTGAC
. S S F L G S R C R K A L N R N P K G S P R F R A . R G K P A N V 270

CGGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATG
CGCTCTTTCCCTTCCCTTCTTTCGCTTTCCTCGCCCGCGATCCCGCGACCGTTACATCGCCAGTGCACGCGCATTGGTGGTGTGGGCGGCGCAATTAC
A R K E G K K A K G A G A R A L A S V A V T L R V T T T P A A L N 280

CGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTGTTATTTTCTAAATACATTCAAAATATGTATCCGCTCAT
GCGGCGATGTCGCGCAGTCCACCGTGAAAAGCCCTTTACACGCGCCTTGGGGATAAACAATAAAAGATTATGTAAAGTTATACATAGGCGAGTA
A P L Q G A S G G T F R G N V R G T P I C L F F . I H S N M Y P L M 290

GAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCTGAGGCGGAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAA
CTCTGTTATTGGGACTATTACGAAGTTATTATACTTTTCTTCTCAGGACTCCGCCCTTCTTGGTGCACCTTACACACAGTCAATCCACACCTT
R Q . P . . M L Q . Y . K R K S P E A E R T S C G M C V S . G V E 300

AGTCCCCAGGCTCCCCAGCAGGCAAGATGCAAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAAGS
TCAGGGTCCGAGGGGTCGCTGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTCAGGGGTCCGAGGGGTGCTCCGCTTC
S P Q A P Q Q A E V C K A C I S I S Q Q P G V E S P Q A P Q Q A E 310

TATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCATCCCGCCCTAACTCCGCCAGTTCGCCCATTTCTCCGCC
ATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTAGGGCGGGTAGGGCGGGGATTAGGGCGGGTCAAGGCGGGTAAGAGGCGGG
V C K A C I S I S Q Q P . S R P . L R P S R P . L R P V P P I L R P 320

CATGGCTGACTAATTTTTTTTATTTATGACAGAGGCCGAGGCCGCTCGGCCCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCT
GTACCGACTGATTAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCGAGAGCTCGATAAGGTCTTCATCACTCTCCGAAAAAACCTCCGATCCGA
M A D . F F L F M Q R P R P P R P L S Y S R S S E E A F L E A . A 330

TTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGAGGTTCTCCGGCCGCTTGGGTGSAGAGGCTAT
AAACGTTTCTAGCTAGTTCTGTGCTTACTCTAGCAAAGCGTACTAATTGTTCTACCTAACGTGCGTCAAGAGGCGGGCAACCCACCTCTCCGATA
F A K I D O E T G . G S F R M I E Q D G L H A G S P A A V V E R L 340

TCGGCTATGACTGGGCACAACAGACAATCGGCTGCTGTATGCCCGGTGTTCCGGCTGTGACGCGAGSGGCGCCCGGTCTTTTTGTCAAGACCGACCT
AGCGGATACTGACCCGTGTGTGCTGTAGCCGACGAGACTACGGCGGCAAGGCGGACAGTCCGCTCCCGCGGGCAAGAAAAACAGTTCTGGCTGGG
F G Y D W A O O T I G C S D A A V F R L S A O G R P V L F V K T D L 350

GTCCGGTGCCCTGAATGAAGTGAAGACGAGGCGAGCGGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGGCGAGCTGTGCTGACGTTGTCTACTGA
CAGGCCACGGGACTTACTTACGTTCTGCTCCGTCGCGCGGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTGCACACGAGCTGCAACAGTGACT
S G A L N E L Q D E A A R L S V L A T T G V P C A A V L D V V T E 360

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 5

GCGGGAAGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATSCAA 370
CGCCCTTCCCTGACCGACGATAACCCGCTTACGGCCCCGTCTAGAGGACAGTAGAGTGGAAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTT
A G R D V L L L G E V P G Q D L L S S H L A P A E K V S I M A D A
TGCGGGGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGCCGGTCTTGT 380
ACGCGCGGACGTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCGGCCAGAACA
M R R L H T L D P A T C P F D H Q A K H R I E R A R T R M E A G L V
CGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTG 390
GCTAGTCTCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGGGTCCGGCTTGACAAGCGGTCGAGTTCCGCTCGTACGGGCTGCGGCTCTTAGAGCAGCAC
D Q D D L D E E H Q G L A P A E L F A R L K A S M P D G E D L V V
ACCATGGCGATGCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGG 400
TGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGCGGAAAAGACCTAAGTAGCTGACACCGGGCGACCCACACCGCTGGCGATACTCC
T H G D A C L P N I M V E N G R F S G F I D C G R L G V A D R Y Q
ACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCTGCTGCTTTACGGTATCGCCGCTCCCGATTTCGACGG 410
TGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCGCGCTTACCCGACTGGCGAAGGAGCAGCAATGCCATAGCGGCGAGGGCTAAGCGTCGC
D I A L A T R D I A E E L G G E V A D R F L V L Y G I A A P D S Q R
CATCGCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGAGATTTCTG 420
GTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCTGCTCGCGGTTGGACGGTAGTGCTCTAAAGC
I A F Y R L L D E F F . A G L V G S K . P T K R R P T C H H E I S
ATTCCACCGCGCCTTCTATGAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT 430
TAAGTGGCGGCGGAAGATATTTCCAACCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGTACGACTCAAGAA
I P P P P S M K G V A S E S F S G T P A G . S S S A G I S C V S S
CGCCACCTTAGGGGAGGCTAACTGAACACGGAAGGAGACAATACCGGAAGGAACCGCGCTATGACGGCAATAAAAGACAGAATAAAACGCAAGGT 440
GCGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTTATGGCTTCTTGGGCGGATACTGCCGTTATTTTCTGCTTATTTTGGCTGCCA
S P T L G G G . L K H G R R Q Y R K E P A L . R Q . K D R I K R T V
GTTGGGTGCTTTGTTTATAAACGCGGGGTTCCGTTCCAGGGCTGGCACTCTGTGATACCCACCGAGACCCATTGGGGCCAATACGCCCCGCTTTCTT 450
CAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCCGACCGTGAGACAGCTATGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAGAA
L G R L F I N A G F G P R A G T L S I P H R D P I G A N T P A F L
CCTTTTCCCCACCCACCCCAAGTTCCGGTGAAGGCCAGGGCTCGCAGCCAACGTCGGGGCGGACGGCCCTGCCATAGCCTCAGGTACTCATATAT 460
GGAAAAGGGTGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAGCGTGGTTGACGCCCGCGCTCCGGGACGGTATCGGAGTCAATGAGTATATA
P F P H P T P Q V R V K A Q G S Q P T S G R Q A L P . P Q V T H I
ACTTTAGATTGATTTAAACTTCATTTTAAATTTAAAGGATCTAGGTGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTCTG 470
TGAAATCTAACTAAATTTGAAGTAAATTTAAATTTTCTAGATCCACTTCTAGGAAAAACATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGC
Y F R L I . N F I F N L K G S R . R S F L I I S . P K S L N V S F R
TTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAGGATCTTCTTGAGATCCTTTTTTCTGCGGTAATCTGCTGCTTGCAACAAAAAACCAACCGC 480
AAGGTGACTCGCAGCTCGGGGATCTTTTCTAGTTTCTAGAGAAGCTCTAGGAAAAAGACGCGCATTAGACGACGAACGTTTGTTTTTTGGTGGG
S T E R O T P . K R S K D L L E I L F F C A . S A A C K Q K N H R

Tuesday, 18 November 1997 11:47
fig 33 pEGFPxba (1 > 5447) Site and Sequence

Page 6

TACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAC TGGCTTCAGCAGAGCGCAGATACCAAACTAGTCCTTCTAGT
ATGGTCGCCACCAACAACCGCCTAGTTCTCGATGGTTGAGAAAAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTATGACAGGAAGATCA 500
Y Q R V F V C R I K S Y Q L F F R R . L A S A E R R Y Q I L S F .
GTAGCGGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACC GCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAG
CATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATT 500
C S R S . A T T S R T L . H R L H T S L C . S C Y Q V L L P V A I S
TCGTGCTTACC GGGTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGCTGAACGGGGGTTTCGTGCACACAGCCAGCTTGGAGCGAA
AGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTTCGGGTCGAACCTCGCTT 510
R V L P G V T Q D D S Y R I R R S G R A E R G V R A H S P A V S E
CGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCGAAGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGT 520
GCTGGATGTGGCTTGACTCTATGGATGTGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCCCTCTTCCGCTGTCCATAGGCCATTTCGCCGTCCCA
R P T P N . D T Y S V S Y E K A P R F P K G E R R T G I R . A A G
CGGAACAGGAGAGCGCAGGAGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGCTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTG 530
GCCTTGCTCTTCGCGTGTCTCTCGAAGGTCCCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAAC
S E Q E S A R G S F Q G E T P G I F I V L S G F A T S D L S V D F C
TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCTTTTACGGTTCTTGCCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTT 540
ACTACGAGCAGTCCCCCGCTCGGATACCTTTTTCGGTCTGTTGCGCGGAAAAATGCCAAGGACCGGAAAAACGACCGGAAAAACGAGTGTACAAGAAAG
D A R Q G G G A Y G K T P A T R P F Y G S V P F A G L L L T C S F
CTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT 5447
GACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
L R Y P L I L W I T V L P P C I

Tuesday, 18 November 1997 11:48

fig 34 pLM4 (1 > 10070) Site and Sequence

Enzymes : 100 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

16p

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCAAGCCATATATGGAGTTCGCGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC 100
pCMV
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T
CCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGT
GGGTGCTGGGGCGGGTAAGTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTCATAAATGCCA 200
pCMV
A Q R P P P I D V N N D V C S H S N A N R D F . P L T S M G G V F T V
AACTGCCCCTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTGAT 300
pCMV
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V
CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACCTACGCCAAAACCGTCATGTAGTTACCGGCACCT 400
pCMV
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A W
TAGCGGTTTGACTCACGGGGATTTCAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTGGCACCAAAATCAACGGGACTTTCCAAATGTCGTA
ATCGCCAACTGAGTGCCCTAAAGGTTAGAGGTTGGGTAAGTGCAGTTACCTCAAAACAAACCGTGGTTTGTAGTTGCCCTGAAAGGTTTACAGCAT 500
pCMV
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .
ACAACTCGCCCCATTGACGCAATGGGCGGTAGGCGGTACGGTGGGAGGTCATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTTGAGCGGGGTAAGTGCCTTTACCGCCATCCGCACATGCCACCTCCAGATATATTCGTCGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT 600
pCMV
Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L
CCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTGCAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCG
GGCAGCGGTGGTACCACTCGTTCCTCGACAAGTGGCCCCACCGGGTAGGACAGCTCGACCTGCCGCTGCATTTGCCGGTGTTCAGTCCG 700
EGFP
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S
TGTCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTTCATCTGCACCACCGGCAAGCTGCCGCTGCCCTGGCCACCTTCGTGAC
ACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTTCAGTGGGACTTCAAGTAGACGTGGTGGCCGTTTCGACGGGCACGGGACCGGTGGGAGCACTG 800
EGFP
V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T
CAECCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCATGAAGCAGCAGCTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCGTGCTGAAGAAGTTCAGGCGGTACGGGCTTCGATGCAGGTCTCT 900
EGFP
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 1

CGCACCATCTTCTTCAAGGACGACGGCAAC TACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC 1000

EGFP

R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC TACAAGCCACAACGCTCTATATCATGGCCGACAAAGCAGAAGAACGGCATCAA
TGAAGTTCCTCCTGCCGTTGTAGGACCCCGTTCGACCTCATGTTGATGTTGTCGGTGTTCAGATATAGTACCGGCTGTTCTGCTCTCTTCCCGTAGTT 1100

EGFP

D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCCTGCGCTCGCACGTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC 1200

EGFP

V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCGCGCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTGAGCGGGGACTCGTTTCTGGGGTGTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACTGGCGGCGGCGCT 1300

EGFP

P D N H Y L S T Q S A L S K D P N E K R D H M V L L E F V T A A G

TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATCTGCAGTCGATAAGCTTGATATCGAATTCCTGCAGCC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGGCTGAGTCTAGAGCTCGAGTTCGAAGCTTAAGACGTCAGCTATTCTGAACATAGCTTAAGGACGTCG 1400

EGFP

I T L G M D E L Y K S G L R S R A Q A S N S A V D K L D I E F L Q P

CCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGC
GGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTCTCTTTCTCTGTCACGCTTACAGGACCTAGAGGCGCTCTGGACCTTCTCTGGTACAGGTG 1500

Insert pLM1

ORF pLM1

L L F S D M L D P E S O R K R T V Q N V L D L R Q N L E E T M S S

CTGCGAGGGTCCAGGTGACTCACAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGAGCGTGTCAGGCTCTCCAACCGCTCG
GACGCTCCAGGGTCCACTGAGTGTGAGGGGACCTCTAC TGGACGATGCTGTCGCTACTACGGTGGGTGCGTCCGACAGGTCCGAGAGGTTGGCGAGCA 1600

Insert pLM1

ORF pLM1

L R G S O V T H S S L E M T C Y D S D D A N P R S V S S L S N R S

CCCCCTGTGTCATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCCCTG
GGGAGACAGTACCGGATACCGGTGAGGTGAGGCGCGGACGTCGACCACTGCGCGGGAGACACCCACCTCGACGCGGAGCTCCCCGCGGGGAG 1700

Insert pLM1

ORF pLM1

S P L S W R Y G O S S P R L Q A G D A P S V G G S C R S E G T P A V

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 3

GTACATGCACGGCGAACGGGCCACTAC TCCACACCATGCCATGCGCA3CCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCGAATCCCTG
CATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGTGGTACGGGTACGGCTCGGGGTCTGTCGAGTCGGTATAGAGGGCGGACCTCGACCAAGCTTAGGGAC

insert pLM1

ORF pLM1

Y M H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L

GACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATG
CTGAGCCTACTCCACCTGGAGTTCAGGCCGATGTACTCGCTGTCTACGGAGTACCGTTCGGTACTGCCCTCC TACTACTGTAGTGTATGGCCGACCTAC

insert pLM1

ORF pLM1

D S D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D

AAAGCAGCTCCATCAGTAGTGACTCAGCGATGCCCTCAGACAATCTCAGTTCAGAAGAATCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCC
TTTCGTCGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGG

insert pLM1

ORF pLM1

E S S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P

CACTGCTTCTCGCAGGAACCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGCTGAGCTGGTTTAGTGAATCAGAGGAG
GTGACGAAGAGCGTCTTGAGTTGTTATCAGGATGCGTGTCTGAGTCTCTTCGGAGTGACCGTCTTTCACCCGACTCGACCAAACTCACTTAGTCTCTC

insert pLM1

ORF pLM1

T A S R R N S T I V L R T D S E K R S L A E S G L S V F S E S E E

AAAGCCCCATAAAAACTGGAGTACGACAGTGGTAGCC TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCAT
FTTCGGGGATTTTGTACCTCATGCTGTCAACATCGGACTTCTACCTTGGACCTGAAGATTACCGCCTCCCTCGCCGGACTCTCGACACTACTAAGTA

insert pLM1

ORF pLM1

K A P K K L E Y D S G S L K M E P G T S K V R R E R P E S C D D S

CCAAGGGTGGAGAAGTGA AAAAGCCATCAGCCTGGGCCACCTGGTTCCTTGAAGAAGGGCAAGACCCACC TG TGGCTGTAAC TCCCCATCACTCA
GGTTCACCTCTTGACTTTTTCGGTAGTTCGGACCCGGTGGGACCAAGGGACTTCTTCCCGTTC TGGGGTGGACACCGACATTGAAGGGGGTAGTGAGT

insert pLM1

ORF pLM1

S K G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

CACAGCCCAGAGTGGCCCTCAAAGTCGCAGGCAAACTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAACTAGGGCTCCAAACGC TCCTCC
GTGTGCGGTCTCACGGGAGTTTCAGCGTCCGTTTGACTCCCGTTTCGATGTCTGTTCCCATTCGAACGTCACATTCATTATGACCCGAGGTTGCGAGGAGG 340

insert pLM1

ORF pLM1

T A Q S A L K V A G K P E G K A T D K G K L A V K N T G L Q R S S

TCTGATGCTGGTGGGACCGCCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCCTC
AGACTACGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGAGCCCGTAACGAGCGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGAGGAGG 350

insert pLM1

ORF pLM1

S D A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P

CTGCCACAGGCACAGCCACTGTATGCAAAGTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGG
GACGGTGTCGTTGCGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCGTTCTAGGTCCTCAGGAGTCCGTAGGGACAGTTTCGGTCATTTACCCGC 360

insert pLM1

ORF pLM1

P A T G T A T V M Q T G G S A T L S K I Q K S S G I P V K P V N G R

CAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGCCAGGATTCTTGGCTCCTGGAGCCGTTCTAACATCCAGTACCGCAGCCTGCCCGGCCAGCCAAAG
GTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCGGTCTTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTCATGGCGTCGGACGGGECGGTTCGGTTC 370

insert pLM1

ORF pLM1

K T S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A I

TCAAGTTCATGAGCGTGACCGGCGGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCTCAGCACCAAGCAGGAGGCCCTTACGC
AGTTCAAGATACTCGCACTGGCCGCCGCCACCCTGGAGCGGGACACTCGTCGTCGTAAC TGGGTCAGAGGAGTCGTGGTTTCGTCCTCCGGAATGCC 380

insert pLM1

ORF pLM1

S S S M S V T G G R G G P R P V S S S I D P S L L S T K Q G G L T

CTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGACCCTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGC
GAAGGTCGACTTCCTCGGATGGTTCCATCGGTCAACCGCC TGGTGAGGTCGGGACAGT TAGTCTGTCTAGCCCTTTTCCTCTCCGGTTTCGGTTCGG 390

insert pLM1

ORF pLM1

P S R L K E P T K V A S G R T T P A P V N Q T D R E K E K A X A I A

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 4

AGTGGCCTTGGACTCAGACAACATCTCC TTGAAGAGTATTGGCTCCCCAGAGAGTACTCCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCA
TCACCGGAACCTGAGTCTGTTGTAGAGGAACCTTCATAACCGAGGGGTCTCTCATGAGGGTTC TTGGTTCGTTCCGGTGGGGTGTCGGTGGTTCGACCGT 300

insert pLM1

ORF pLM1

V A L D S D N I S L K S I G S P E S T P K N Q A S H P T A T K L A

GAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGAGCTTTGTCAAACCAACCTCACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTAC
CTCGACGGTGGTTGGGAGAGTCCCGGTGTCGCTTCTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACTGTTCAGTTGAGGTTGTCAGACCTAGATG 310

insert pLM1

ORF pLM1

E L P P T P L R A T A K S F V K P P S L A N L D K V N S N S L D L

CATCATCCAGTGATACCAACCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCTGCTTACCCCCAGTCC
GTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTTCGAGTCGTAGACCCCGGGAGAGGGAAGGACGAAGTGGGGGTCAGG 320

insert pLM1

ORF pLM1

P S S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P

GGCACCATCTCAATATTAAGTCAAGCCAGCTTCTCCCAGGGCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAATC
CCGTGGGTAGGAGTTATAATTGAGTCGGTCGAAGAGGGTCCCGGACCTCGATTACTACCAAAGTCACACGGTTTCTCTGGGCGTACATGGGGTTTGAG 330

insert pLM1

ORF pLM1

A P I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L

TCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTG
AGTCCGGACGTGCTCTCGTACCTCAGGGAGGTCTACGGTTAC TCGGAGGGGTACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGAC 340

insert pLM1

ORF pLM1

S G L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P

CTGCTCCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAA
GACGAGGGTGTCTTCTCTGCTTCTCGACTGAACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCTATTAGTCGCCCTAGCCTTGTGAGAAGGGTT 350

insert pLM1

ORF pLM1

A A P T E E E T E E L T V S G S P R A G O L D S N O R D R N T L P I

Tuesday, 18 November 1997 11:48
fig 34 pLM4 (1 > 10070) Site and Sequence

Page 6

GAAAGGCTCAGGTACCAGCTTCAGTCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATTTGGTGGGCTGCCGTAATCCGATGACCAGTCAGAG
CTTTCCGAGTCCATGGTCAAGTCAGGGTCTCTCTGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCGACGGACTTAGGCTACTGGTCAGTCTC

3600

insert pLM1

ORF pLM1

K G L R Y Q L Q S Q E E T K E R R H S H T I G G L P E S D D Q S E

CTGCCTTCTCCCTGCACCTTCCCATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCAC TGCGGCCACCCAGCCAAGAAATCACCCGCT
GACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTGAATGGTGTATCACTCAGGGTGACGCCGGTGGTGGGTTCTTAGTGGGGCA

3700

insert pLM1

ORF pLM1

L P S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R

CCAACAGCATCCCCACCCACGAGGGGCTTCGAGCTGTACAGCGGCTCCCAAATGGGGAGCACCC TGTCCTGGCCGAGAGACCCAAGGGAATGATTCC
GGTTGTCGTAGGGGTGGGTGCTCCGCCGGAAGCTCGACATGTGCGCCGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGC

3800

insert pLM1

ORF pLM1

S N S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R

GTCAGGATCCTTCGAGACCCACGGACGATGTTACGGCTCAGTGCTGTCCCTGGCCTCCAGTGCCCTCCACCTACTCCTCAGCTGAGGAGAGGATG
CAGTCTTAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTCACGACAGGGACCGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTAC

3900

insert pLM1

ORF pLM1

S G S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M

CAATCTGAGCAAAATCCGGAAGCTTCGTAGGGAAC TGAATCATCCAGGAAAAGTGCCACCTTGACGTC TCAGCTCTCTGCCAATGCTAATCTGGTGG
GTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTCAACGGTGGAAC TGCAGAGTCGAAAGACGGTTACGATTAGACCAAC

4000

insert pLM1

ORF pLM1

D S E O I R K L R R E L E S S Q E K V A T L T S O L S A N A N L V

CTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCTGCGACACC TGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTGCGAGAAAC
GACGAAAAC TCGTCTCGGACCAC TTACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCTGTGACTCGACGACCTAAACGCTCTTTG

4100

insert pLM1

ORF pLM1

A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L D L R E T

Tuesday, 18 November 1997 13:56
fig 53 pLM6 (1 > 4947) Site and Sequence

Page 3

AGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCCAACCAACTGGATCA
TCTTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTCCCTTACTTCGAATGCTGTAGGCGAACC TCCGGGAGTTGAGACGGGTGCTTGACCTAGT

U3 stuk

ORF

K K E V S E L R S E L V E X E M K L T D I R L E A L N S A H Q L D Q

GCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACT
CGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCTTACTGGCTGACTTCCATCGGGGTCGGGGAGTAGTCCGAGGTGA

U3 stuk

ORF

L R E T M H N M Q L E V D L L K A E N D R L K V A P G P S S G S T

CCAGGGCAGGTCCTGGATCATCTGCATTATCTTCCCCACGCCCTCCCTAGGCCCTGGCACTCACCCATTCTTCGGCCCCAGTCTTGCAGACACAGACC
GGTCCCGTCCAGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAGGAAGCCGGGTCAGAACGCTGTGTCTGG

U3 stuk

ORF

P G Q V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D

TGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAGGGGACTT
ACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCTCTTCACTGGGAGGCCACCACCACTCTACGGGGGCGTCTGTAGTAGTTTCCCCTGAA

U3 stuk

ORF

L S P M D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L

GAAGCAGCAGGAATTCTTCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTCCAAGTGTTCAGGACTATAT
CTTCGTCTCTTAAGAAGGACCCGACATCGTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTCAAGTTCTTGATATAA

U3 stuk

ORF

K Q Q E F F L G C S K V S G K V D W K M L D E A V F Q V F K D Y I

TCTAAATGGACCCAGCCTCTACCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCCGS
AGATTTTACCTGGGTCGGAGATGGGACCTGATTCGTGACTCAGGTAGGTACCGATGTCTAGTTCGGTGCACTTTGCTCACAACCTACGCTCTCGGGGGG

U3 stuk

ORF

S K M D P A S T L G L S T E S I H G Y S I S H V K R V L D A E P P

Tuesday, 18 November 1997 13:57
fig 53 pLM8 (1 > 4947) Site and Sequence

Page 4

AGATGCCCTCCTTGCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCC
TCTACGGAGGAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGCAGCTGTCGGACCACAAGCTCTGCGACTAGGS 3000

U3 stuk

ORF

E M P P C R R G V N N I S V S L K G L K E K C V D S L V F E T L I P

CAAGCCGATGATGCAGCACTACATAAGCCTCTGTGAAGCACCAGGCGCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGC
GTTCCGGCTACTACGTCGTGATGATTTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTCTTGGATGGACTGGTTAGCG 3100

U3 stuk

ORF

K P M M O H Y I S L L L K H R R L V L S G P S G T G K T Y L T N R

TTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGT
AACCAGGCTCATGGACCACCTCGCGAGACCGGCACCTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTAGACGTTGACA 3200

U3 stuk

ORF

L A E Y L V E R S G R E V T E G I V S T F N M H O O S C K D L O L

ATCTTTCCAACCTAGCCAACAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAG
TAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCCTAAGATAACCTACTGGACTCACTTCGTCCGAGGTTAGTC 3300

U3 stuk

ORF

Y L S N L A N O I D R E T G I G D V P L V I L L D D L S E A G S I S

TGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCAACCATGGCTTG
ACTCAACCAGTTACCCCGGAGTGGACGTTTCATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTAC TGTGGGTGGTACCGAAC 3400

U3 stuk

ORF

E L V N G A L T C K Y H K C P Y I I G T T N O P V K M T P N H G L

CACCTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCG
GTGAACCTCGAAGTCTACAACGGAAGAGGTTGTTGCACCTCGGTTCGGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCGTCTCGC 3500

U3 stuk

ORF

H L S F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S

Tuesday, 18 November 1997 13:57
fig 53 pLM8 (1 > 4947) Site and Sequence

Page 5

ACATCAATGCCAACGAAGGAGAGCTGCTCGGGTGCCTGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGA
TGTAGTTACGGTTGTTCTCTCGACGAAGCCACGAGCTGACCCATGGGTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGCTGCTGTSAGTC 360

U3 stuk

ORF

D I N A N K E E L L R V L D V V P K L V Y H L H T F L E K H S T S D

CTTCCTCATCGGCCCTTGCTTCTTCTGCTGCTCCATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACCTATCATTCCCTAT
GAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCTTGGACCAAGTAACGGACACCTTGTGAGATAGTAAGGGATA 370

U3 stuk

ORF

F L I G P C F F L S C P I G I E D F R T V F I D L V N N S I I P Y

CTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCTTGCCATCAG
GATGTCTTCTCGGTTCCTACCTATTTCAGGTACCTGTCTTTCGACGAACCTCTCGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTC 380

U3 stuk

ORF

L O E G A K D G I K V H G O K A A W E D P V E V V R D T L P W P S

CCCAACAAGACCAATCAAGCTGTACCACTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCAC
GGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCTGAACGGAGTGGAGGGCTCCATCTGTGACGTTTCTGTCTGTS 390

U3 stuk

ORF

A Q Q D O S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T

CCCAAGTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCTGGAC
GGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTTGAAGTCTTCGACGGTTGATGTAACCTCAGAGGTCTAGCTCTTTGGTAGGACCTG 400

U3 stuk

ORF

P S S L D S D P L M A M L L K L O E A A N Y I E S P D R E T I L D

CCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTACCCCGGACAGCAGAACGCTGGCATCAGCTATCTTAGCTCTCTCTCCCTCTCTCC
GGGTTGGAAGTCCGTTGTGAAATTCCCAAGCGTTAGTGACAGTGGGGGCTGTGCTCTTGCACCGTAGTCGATAGAATCGAAGGAGAGGGAGAGG 410

U3 stuk

ORF

P N L O A T L G F G N H C H P R T A E R V H O L S L L L S P L

Page 6

[illegible]

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Enzymes : 72 of 148 enzymes (Filtered)

Settings : Circular, Certain Sites Only, Standard Genetic Code

Page 1

GTGGCACATTTTCGGGAAATGTGCGGGAACCCCTATTGTATTATTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAAATACCCCTGATAAAT 100
CACCGTGAAAAGCCCTTTACACGCGCCTTGGGGATAACAAATAAAAGATTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTA
G G T F R G N V R G T P I C L F F . I H S N M Y P L M R O . P . . M
GCTTCAATAATTTGAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCTTTTTCGCGCATTTTGCTTCTCTGTTTGTCTCAC 200
CGAAGTTATTATAACTTTTCTCTCTACTACTAAGTTGTAAAGGCACAGCGGGAATTAAGGAAAAACGCCGTAAACGGAAGGACAAAAACGAGTG
L Q . Y . K R K S M S I Q H F R V A L I P F F A A F C L P V F A H
CCAGAAACGCTGGTGAAGTAAAGATGCTGAAGTACAGTTGGGTGCACGAGTGGGTACATCGAACTGGATCTCAACAGCGGTAAAGATCCTTGAGAGTT 300
GGTCTTTGGCACCCTTTCAATTTCTACGACTTCTAGTCAACCCACGTCCTACCCCAATGATGCTTACCTAGAGTTGTGCGCATTTCTAGGAACCTCTCAA
P E T L V K V K D A E D O L G A R V G Y I E L D L N S G K I L E S
TTGCCCCGAAGACGTTTTCATGATGAGCACATTTAAAGTTCTGCTATGTGGCGCGGTATTATCCGCTATTGACGCGGGCAAGGCAACTCGGTG 400
AAGCGGGGCTTCTGCAAAAGGTTACTACTCGTGAAATTTCAAGACGATACACCGGCCATAAGGCAATACTGCGGCCGTTCTGTTGAGCCAGC
F R P E E R F P M H S T F K V L L C G A V L S R I D A G O E O L G R
CCGATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGACGTGCTGCC 500
GGCGTATGTGATAAGAGTCTTACTGAACCACTCATGAGTGGTCAGTGCTTTTCTAGTAATGCTACCGTACTGTCATTCTCTTAATACGTACGACGG
R I H Y S Q N D L V E Y S P V T E K H L T D G M T V R E L C S A A
ATAACCATGAGTGATAACACTGCGGCCAATTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTCACACATGGGGATCATGTAA 600
TATTGGTACTCACTATTGTGAGCGCGGTGAATGAAGACTGTGCTAGCTCTCTGGCTTCTCGATTGGCGAAAAACGTGTGTACCCCTAGTACATT
I T M S D N T A A N L L L T T I G G P K E L T A F L H N M G D H V
CTGCGCTTGATCGTTGGGAACCGAGCTGAATGAAGCATACCAACGACGAGCGTGACACACGATGCCGTAGCAATGGCAACAACGTTGCGCAAACT 700
GAGCGGAATAGCAACCTTGGCTCGACTTACCTCGGTATGGTTTGTGCTGCGACTGTGGTGTACGGACATCGTTACCGTTGTTCAACGCGTTTGA
T R L D R V E P E L N E A I P N D E R O T T M P V A M A T T L R K L
ATTAACGCGAATCTACTCTAGCTTCCCGCAACAATTAATAGACTGGATGGAGGCGGATAAGTTGACGAGCACCTTCTGCGCTCGGCCCTTCCG 800
TAATTGACCGCTTGATGAATGAGATCGAAGGCGGTGTAAATATCTGACCTACCTCCGCTATTTCAACGCTCTGGTGAAGACGCGAGCCGGGAAGGC
L T G E L L T L A S R Q O L I D V H E A D K V A G P L L R S A L P
GCTGGCTGGTTTATTGCTGATAAATCTGGAGCGGTGAGCGTGGGTCTCGCGGTATCATGCAAGCTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG 900
CGACCGACCAATAACGACTATTATACCTCGGCCACTCGCACCCAGAGGCCATAGTAACCTCGTGACCCCGGTCTACCAATCGGGAGGGCATAGCATC
A G V F I A D K S G A G E R G S R G I I A A L G P D G K P S R I V
TTATCTACACGAGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAAGTCCCTACCTGATTAAAGCATTGGTAACGTGACACCA 1000
AATAGATGCTGCCCTCAGTCCGTTGATACCTACTTGTCTTATCTGCTAGCGACTCTATCACGGAGTGACTAATTCGTAACCATTGACAGTCTGGT
V I Y T T G S Q A T M D E R N R O I A E I G A S L I K H V . L S D Q
AGTTTACTCATATACTATTAGATTGATTTAAACTTCAATTTTAAATTTAAAGGATCTAGGTAAGATCCTTTTGTATAATCTCATGACCAAAATCCCT 1100
TCAATGAGTATATAGAACTAATAAATTTGAAGTAAATTAATTTTCTAGATCCACTTCTAGGAAAACTATTAGAGTACTGGTTTATAGGGA
V Y S Y I L . I D L K L H F . F K R I . V K I L F D N L M T K I P
TAACGTGAGTTTTCGTTCCACTGACGCTGACACCCGCTAGAAAAAGATCAAGGATCTCTTGAAGATCCTTTTTCGCGGTAATCTGCTGCTTGCAAA 1200
ATTGCACTCAAAAGCAAGGTGACTGCGAGTCTGGGGCATCTTTCTAGTTCTTAGAAGAACTCTAGGAAAAAGACGCGCATTAGACGACGAACGTTT
. R E F S F H . A S D P V E K I K G S S . D P F F L R V I C C L O
CAAAAAACACCGCTACCGAGCGTGGTTGTTTGGCGGATCAAGAGCTACCAACTCTTTTTCGAAGGTAACGGCTTACGACAGCGCAGATACCAAA 1300
GTTTTTTTGGTGGCGATGGTGGCCACCAACACCGGCTAGTTCTCGATGGTGTAGAAAAAAGTCTCCATTGACCGAAGTCGCTCGCGCTCTATGGTTT
T K K P P L P A V V C L P D O E L P T L F P K V T G F S R A Q I P N
TACTGCTCTCTAGTGTAGCGTAGTATAGGCCACCACTTCAAGAACTCTGTGACACCGCTCTATACCTCGCTCTGCTAATCTGTTACCAAGTGGCTGCT 1400
ATGACAGGAAGATCAGATCGGCATCAATCCGCTGGTGAAGTTCTTGAGACATCTGGCGGATGATGGAGCGAGAGGATTAGGACAAATGGTACCGACGA
T V L L V . P . L G H H F K N S V A P P T Y L A L L I L L P V A A

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 2

GCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTGGGGCTGAACGGGGGGTTCGTGCACACAGC
CGGTCAACCGCTATTCAGCACAGAATGGCCCAACCTGAGTTCGTATCAATGGCTATTTCGCGCTGCCAGCCGACTTGCCTCCCAAGCACGTGTGTGG 1500
A S G D K S C L T G L D S R R . L P D K A Q R S G . T G G S C T O

CCAGCTTGGAGCGAACGACCTACACCGAAGTACGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGAGAAAGGCGGACAGGTATCC
GGTCGAACCTCGCTTGGATGTGGCTTGACTCTATGGATGTCGCACTCGGATACCTTTTCGCGGTGCGAAGGGCTTCCCTCTTTCGCGCTGTCCATAGG 1600
P S L E R T T Y T E L R Y L O R E L . E S A T L P E G R K A D R Y P

GGTAAGCGCGAGGGTCGGAACAGGAGAGCGCACGAGGAGCTTCCAGGGGGAACGCCCTGGTATCTTTATAGTCTGTGGGTTCGCCACCTCTGACTT
CCATTCGCGCTGCCAGGCTTGTCTCTCGCGTCTCCCTCGAAGGTCCCTTTGCGGACCATAGAAATATCAGGACAGCCCAAGCGGTGGAGACTGAA 1700
V S G R V G T G E R T R E L P G G N A V Y L Y S P V G F R H L . L

GAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAACACGCCAGCAACGCGGCTTTTACGGTTCCTGGCTTTTGTGGCTTTTG
CTGCGAGCTAAACACTACGAGCAGTCCCGGCTCGGATACCTTTTTCGCGTCTGCGCGGAAATGCCAAGGACCGGAAACGCGGAAAC 1800
E R R F L . C S S G G R S L V K N A S N A A F L R F L A F C V P F

CTCACATGTTCTTCTCGCTTATCCCTGATTCTGTGGATAACCGTATTACCGCTTTGAGTGAGCTGATACCGCTCGCGCAGCGCAACGACGAGCG
GAGTGTACAAGAAAGGACGCAATAGGGACTAAGACACCTATTGGCATAATGGCGGAACTCACTGAGCTATGGGAGCGGCTCGGCTTGTGGCTCGC 1900
A H M F F P A L S P D S V D N R I T A F E . A D T A R R S R T Y E R

CAUCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAACCGCTCTCCCGCGGTGGGCGATTTCATTAATGACGTGGCAGCAGAGTTT
GTCGCTCAGTCACTCGCTCTTTCGCTTCTCGCGGTTATGCGTTTGGCGGAGAGGGGCGCAACCGCTAAGTAATTACGTCGACCGTGTGTCCAAA 2000
S E S V S E E A E E R P I R K P P L P A R V P I H . C S V H D R F

CCGAC TGGAAAGCGGCGAGTGAGCGCAACGCAATTAAATGAGTTAGCTCACTCATTAGGCAACCGGCTTACACTTTATGCTTTCGGCTCGTATGT
GGGCTGACCTTTTCGCGCTCACTCGGCTTGGCTTAATTACACTCAATCGAGTGAGTAATCCGTGGGTCGGAATGTGAATACGAAGGCGGAGCATACA 2100
P D V K A G S E R N A I N V S . L T H . A P Q A L H F H L P A R H

TGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCTCACTAAAGGAACAAAGCT
ACACACCTTAACACTCGCTTATGTTAAAGTGTGCTTTGTCGATACCTGAGTAAATGCGGTTTCGCGGTTAATTGGGAGTGATTCCCTTGTTTTCGA 2200
L C G I V S G . O F H T G N S Y D H D Y A K R A I N P H . R E Q K L

GGGTACCGGGCCCCCTCGAGGTGACGATATCGATAAGCTTGAATCGAATCTCGAGCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCAG
CCATATGGCCCGGGGGGAGCTCCAGCTGCCATAGCTATTGCAACTATAGCTTAAGGAGCTCGGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGGTC 2300

insert pLM1

ORF pLM1

G T G P P L E V D G I D K L D I E F L Q P L L F S Q M L D P E S O

AGAAAGAGGACAGTGCAGAAATGCTCTGGATCTCCGCGAGAACC TGGAAAGAGCATGTCCAGCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGAGA 2400
TCITTCCTCTGTCAGCTCTTACAGGACCTAGAGGCGCTTTGGACCTTCTCTGGTACAGGTGCGAGCTCCAGGGTCCACTGAGTGTGAGGGACCTCT

insert pLM1

ORF pLM1

R K R T V O N V L D L R Q N L E E T M S S L R G S Q V T H S S L E

TGACCTGCTACGACAGCGATGATGCCAACCCACGAGCGTGTCAGCTCTCCAACCGCTCGTCCCTCTGTATGGCGCTATGGCCAGTCCAGTCCGCG 2500
ACTGGACGATGCTGCTGCTACTACGGTTGGGTGCGTCGCACAGGTGCGAGAGGTGGCGAGCAGGGAGACAGTACCGGATACCGGTCAGGTACGGCGC

insert pLM1

ORF pLM1

H T C Y D S D D A N P R S V S S L S N R S S P L S V R Y G C S S P R

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 3

GC TGCAGGCTGGTGACGCGCCCTCTGTGGG TGGGAGC TGCCGCTCGGAGGGGACGCCCGCTGGTACATGCACGGCGAACGGGCCAC TACTCCACACC 2800
CGACGTCCGACCCTGCGCGGGAGACACCCACCTCGACGGCGAGGCTCCCTCGGGGGGACCATGTACGTGCCGCTTGGCCGGGTGATGAGGGGTGG
-----insert pLM1-----
-----ORF pLM1-----
L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H T
ATGCCCATGCCAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACATGA 2700
TACGGGTACGGCTCGGGTCTGTCGAGTCGGTATAGAGGGCGGACCTCGACAGCTTAGGGACCTGAGGCTTACTCCACCTGGAGTTCAGGCCGATGTACT
-----insert pLM1-----
-----ORF pLM1-----
H P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y M
GCGACAGTGACCTCATGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGACTCAGCGATGCCCTC 2800
CGCTGTCACTGGAGTACCGGTTCTGGTACTGCCCTCTACTACTGTAGTGATGGCCGACCTACTTTCTGTCGAGGTAGTCATCACCTGAGTCGCTACGGAG
-----insert pLM1-----
-----ORF pLM1-----
S D S D L M G K T M T E D D D I T T G V D E S S S I S S G L S D A S
AGACAATCTCAGTTCAGAAGAAATCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGEAGGAACCAACAATAGTGCTACGG 2900
TCTGTAGAGTCAAGTCTTCTTAAGTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGACGAAGAGCGTCTTGAGTTGTTATCAGCATGCG
-----insert pLM1-----
-----ORF pLM1-----
D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L R
ACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTAGTGAAATCAGAGGAGAAAGCCCCATAAAAACTGGAGTACGACAGTGGTAGGCC 3000
TGCTGAGTCTCTTCGGAGTGACCGCTCTTCAACCGGACTCGACCAAACTACTAGTCTCTCTTTCGGGGATTTTTCACCTCATGCTGTCACCATCGG
-----insert pLM1-----
-----ORF pLM1-----
T D S E K R S L A E S G L S V F S E S E E K A P K K L E Y D S G S
TGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCATCAAGGGTGGAGAATGAAAAAGCCCATCAGCCTGGG 3100
ACTTCTACCTTGGACCCTGAAGATTCACCGCTCCCTCGCGGACTCTCGACACTACTAAGTAGGTTCACCTCTTGAC TTTTCGGGTAGTCGGACCC
-----insert pLM1-----
-----ORF pLM1-----
L K M E P G T S K V R R E R P E S C D D S S K G G E L K K P I S L G
CCACCTGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCATCACTCACACGCCAGAGTGCCCTCAAAGTCGAGGCAAACT 3200
GGTGGGACCAAGGACTTCTTCCGTTCTGGGGTGGACACCGACATGAAGGGGGTAGTGAGTGTGTCGGGCTCACGGGAGTTTCAGCGTCCGTTTGA
-----insert pLM1-----
-----ORF pLM1-----
H P G S L K K G K T P P V A V T S P I T H T A Q S A L K V A G K P
GAGGGCAAAGCTACAGACAAGGTAAGCTTGCAGTGAAGAATACTGGGCTCAACGCTCTCTCTGATGCTGGTCGGGACCGCTGAGTGATGCTAAGA 3300
CTCCCGTTTCGATGCTGTTCCTTCCATTCAAGCTCACTTCTTATGACCCGAGGTTCGAGGAGGAGACTACGACAGCCCTGGCGGACTCACTACGATCT
-----insert pLM1-----
-----ORF pLM1-----
E G K A T D K G K L A V K N T G L O R S S S D A G R D R L S D A K

Tuesday, 18 November 1997 13:57

(fig 54 pLM1 (1 > 8285) Site and Sequence

Page 4

AGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCCTCGGCTACAAGAAGCCCTCCCTGCCACAGGCACAGCCACTGTCATGCAAACTGGTGG 3400
TCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTCGGAGGAGGACGGTGTCCGTGTCGGTGACAGTACGTTTGACCACC
insert pLM1
ORF pLM1
K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G G
TTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCAACAGCGCAGAGCCA 3500
AAGTCGGTGAGAGTCGTTCTAGGTCTCAGGAGTCCGTAGGGACAGTTCGGTCATTTACCCGCGTTCGATCGAATCTACAAAGGTTGTCGGCTCTCGGT
insert pLM1
ORF pLM1
S A T L S K I O K S S G I P V K P V N G R K T S L D V S N S A E P
GGATTCTCGGCTCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCCTGCCCCGGCCAGCCAAAGTCAAGTTCTATGAGCGTGACCGCGGGGGGGTGGAC 3600
CCTAAGGACCGAGGACCTCGGGCAAGATTGTAGTTCATGGCTCGGACGGGCGCGGTTCGGTTCAGTCAAGATACTCGCACTGGCCGCGCCCACTG
insert pLM1
ORF pLM1
G F L A P G A R S N I O Y R S L P R P A K S S S H S V T G G R G G
CTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCCTCAGCACCAAGCAGGAGGGCCCTACGCCCTTCAGACTGAAGGAGCC TACCAAGGTAGCCAGTGG 3700
GAGCGGGCACTCGTCTGTAAC TGGGTCAGAGGAGTCG TGGTTCGTCCTCCGGAATCGGGAAGGTC TGACTTCCTCGGATGGTTCATCGGTCAAC
insert pLM1
ORF pLM1
P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S G
GCGGACCACTCCAGCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAAGCAGTGGCTTGGACTCAGACAACATCTCCTTGAAGAGT 3800
CGCC TGGTGAGGTCGGGGACAGTTAGTCTGTCTAGCCCTTTCTCTTCCGGTTTCGGTTCGTCACCGGAACCTGAGTCTGTTGAGAGGAACCTCTCA
insert pLM1
ORF pLM1
P T T P A P V N O T O R E K E K A K A V A L O S D N I S L K S
ATGGCTCCCCAGAGTACTCCCAAGAACAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGCGAAGA 3900
TAACCGAGGGGCTCTCATGAGGGTCTCTGGTTCGTTCCGTTGGGTCGCGTGGTTCGACCGTCTCAGCGGTGGTGGGGAGAGTCCCGGTGTCGCTCT
insert pLM1
ORF pLM1
I G S P E S T P K N O A S H P T A T K L A E L P P T P L R A T A K
GTTTGTCAAAACCCCTCACTAGCCAATCTTGACAAGGTCAACTCCAAAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAGGTCCC 4000
CJAACAGTTTGGTGGGAGTGATCGGTTAGAATGTTCCAGTTGAGGTTGTGACACCTAGATGGTAGTAGGTCACTATGGTGGGTACGAAGTTTCCAGGG
insert pLM1
ORF pLM1
S F V K P P S L A N L O K V N S N S L D L P S S S D T T H A S K V P
AGATCTGATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTACCCCCAGTCCGGCACCCATCTCAATATTAACCTAGCGAGCTTCTCC 4100
CTAGACGTACGATGTTGAGTCGTAGACCCCGGAGAGGGAAGGACGAAGTGGGGTTCAGGCCGTGGTAGGAGTTATAATTGAGTCGGTCAAGAGG
insert pLM1
ORF pLM1
U L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F S

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 5

CAGGGCCCTGGAGCTAATGAGTGGTTCAGTGTGCCAAAAGAGACCCGATGTACCCCAAACCTCAGGGCTGCACAGGAGCATGGAGTCCCTCCAGATGC 4200
GTCCCGGACCTCGATTACTACCAAAGTCACACGGTTCCTCTGGGCGTACATGGGTTTGAGAGTCCGGACGTGTCTCGTACCTCAGGGAGGTCTACG
-----insert pLM1-----
-----ORF pLM1-----
Q G L E L N S G F S V P K E T R M Y P K L S G L H R S M E S L O M
CAATGAGCCTCCCAAGTGCCTTCCCAAGCAGTACTCCCGTCCCAACCCACCTGCTCCCTGCTGCTCCACAGAAGAAGAGACGGAAGAGCTGACTTG 4300
GTACTCGGAGGGGTCACGGAAGGGTCTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTCTGCTTCTCGACTGAAC
-----insert pLM1-----
-----ORF pLM1-----
P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T V
GAGTGAAGCCCAAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAG 4400
CTCACCTTCGGGGTCTCGACCCGTGACCTGTCTATAGTCGCCCTAGCCTGTGAGAAGGGTTCCTTCCCGAGTCCATGGTCGAAGTCAGGGTCCCTCCTC
-----insert pLM1-----
-----ORF pLM1-----
S G S P R A G O L D S N O R O R N T L P K K G L R Y O L Q S Q E E
ACCAAGGAGAGGGCAGATTCCTACCATTTGGTGGGCTGCCGTAATCCGATGACAGCTAGAGCTGCTTCTCCCTGCACTTCCCATGCTCTGAGTG 4500
TGGTTCCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTCAC
-----insert pLM1-----
-----ORF pLM1-----
T K E R R H S H T I G G L P E S D D O S E L P S P P A L P M S L S
CAAAGGGCCCACTTACCAACATAGTGAGTCCCACTGCGGCCACCAACGCAAGAATCACCCTGCAACAGCATCCCAACCCACGAGGGCGGCTTCGAGCT 4600
GTTTCCCGGTTGAATGGTTGTATCATCAGGGTGACCGGTTGGTGGGTTCTTAGTGGGCGAGGTGTCTGAGGGGTGGGTGCTCCCGCGGAAGCTCGA
-----insert pLM1-----
-----ORF pLM1-----
A K G O L T N I V S P T A A T T P R I T R S N S I P T H E A A F E L
GTACAGCGGCTCCCAATGGGAGCACCTGTCTCCGCGAGAGACCAAGGGAATGATTCGGTCAGGATCCTTCCGAGACCCACGAGCAGATGTTTAC 4700
CATGTGCGCGAGGGTTTACCCCTCGTGGGACAGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCTAGGAAGGCTCTGGGGTGCTGCTACAAGTG
-----insert pLM1-----
-----ORF pLM1-----
Y S G S O M G S T L S L A E R P K G M I R S G S F R D P T D D V H
GGCTCAGTGCTGTCCCTGGCCTCCAGTGCCTCTCCACCTACTCTCAGCTGAGGAGAGGATGCAATCTGAGCAAAATCCGGAAGCTTCGTAGGGAAGTGG 4800
CCGAGTCACGACAGGGACCGGAGGTACGAGGAGGTGGATGAGGAGTGCAGTCTCTCTACGTTAGACTCGTTTAGGCTTCGAAGCATCCCTTGACC
-----insert pLM1-----
-----ORF pLM1-----
G S V L S L A S S A S S T Y S S A E E R M O S E O I R K L R R E L
AATCATCCCAAGGAAAGTGGCCACCTTGACGTCTCAGCTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCTGGTGAATATGACATCCCG 4900
TTAGTAGGGTCTCTTTTACCGGTGAAC TGCAGAGTGGAAAGACGGTTACGATTAGACCAACGACGAAACCTGCTCGGACCACTTATATGTAGGGC
-----insert pLM1-----
-----ORF pLM1-----
E S S O E K V A T L T S O L S A N A N L V A A F E O S L Y N M T S R

Page 6

[illegible]

Tuesday, 18 November 1997 13:57
fig 54 pLM1 (1 > 8285) Site and Sequence

Page 7

TCACCCATTCCTTCGGCCCCAGCTTCGACACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCCTCGGGTGGT
AGTGGGTAAGGAAGCCGGGGTCAGAACGCTCTGTGCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAGGTTTCCCTCCTCACTGGGAGGCCACCA 5800

insert pLM1

ORF pLM1

L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V V

GGTGAGGATGCCCCGAGCACATCATCAAGGGGACTTGAAGCAGCAGGAATTCCTTGGGCTGTAGCAAGGTCACTGGAAAAGTTGACTGGAAGATG
CCACTCCTACGGGGGCGTGTGTAGTAGTTTCCCTGAACTTCGTCGTCTTAAGAAGGACCCGACATCGTTCAGTCACCTTTCAACTGACCTTCTAC 5900

insert pLM1

ORF pLM1

V R M P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D V K H

CTGGATGAAGCTGTTTCCAAGTGTCAAGGACTATATTCTAAATGGACCCAGCCCTTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCA
GACCTACTTCGACAAAAGGTTCAAGTTCTGATATAAAGATTTACCTGGGTCGGAGATGGGACCTGATTCTGTGACTCAGGTAGGTACCGATGTCTG 6000

insert pLM1

ORF pLM1

L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

TCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCGAGATGCCCTCTTGGCGTCGAGGTGTAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGA
AGTCTGGTGCACTTTGCTCACAACCTACGTC TCGGGGGGCTCTACGGAGGAAACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCT 6100

insert pLM1

ORF pLM1

I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L K E

GAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGACGACTACATAAGCCTCTGCTGAAGCACGGGCGCTCGTCTCTCG
CTTTACGCACTGTCCGACCACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTTCGGAGGACGACTTCGTGGCCCGGAGCAGGAGAGC 6200

insert pLM1

ORF pLM1

K C V D S L V F E T L I P K P M H O H Y I S L L L K H R R L V L S

GGCCCCAGCGGCACGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTSAGGTACAGAGGGCATCGTCAGCACCT
CCUGGGTCGCGTGCCGCTTCTGGATGGACTGGTTAGCGAAGCCGGCTCATGGAACACCTCGCGAGACCGGCATCCAGTGTCTCCCGTAGCAGTCGTGGA 6300

insert pLM1

ORF pLM1

G P S G I G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

TAAACATGCACCAAGCTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTGGT
AGTTGTACGTGGTGTGAGAACGTTCTAGACGTTGACATAGAAAGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCA 6400

insert pLM1

ORF pLM1

F N M H O O S C K D L O L Y L S N L A N O I D R E T G I G D V P L V

GATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTATAGGTACC
CTAAGATAACCTACTGGACTCACTTCGTCGAGGTAGTCACTCAACCAGTTACCCCGGAGTGGACGTTTCATAGTATTACAGGGATATAATATCCATGG 6500

insert pLM1

ORF pLM1

I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G T

Page 4

[illegible]

Tuesday, 18 November 1997 13:57

fig 54 pLM1 (1 > 8285) Site and Sequence

Page 1

AUGAGGGACAGGTTCTTGTGCTGTACCTTTGAGAACCTCC TAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTTGTCCTCCCTTAAACACATTTACTGGC
TCTTCCCTGTCCAAGAACCACGACATGGAACTCTTGAAGGATCTTCTTACCACCCACCGCAAACCTTGAACACGGGGGATTGTGTAAATGACCG 7400

— insert pLM1 —
G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T G

CTCC TC TAATGACTTTGGGAAAAGATGATTCTGGGTCTTCCCTTGACTTCTTGTTC AATTACAACTCTGGGCTTCTGGGGAGGGGTT CAGAAAA
GAGGAGATTACTGAAACCCCTTTTCTACTAAGACCAGAAAGGGAACGAAGAACAAGTTAATGTTTGAGGACCGAAAGACCCCTCCCAAGTCTTTT 7500

— insert pLM1 —
L L . . L V G K D D S G S F P . L L V S I T N S V A F V G G V Q K

CATCAAACTCTGCAGCAGTTCCCGGAATTCAGCTTGGACTTAACCAAGGCTGAACCTTGCTCAAAAGAACCGAATTCAGCACACTGGCGGCGTTACT
GTAGTTTGTGACGTCGTAAGGGGCTTAAGTCGAACCTGAATTTGGTCCGACTGAACGAGTTTCTTCGGCTTAAGGTCGTGTACCGCGGCAATGA 7600

— insert pLM1 —
T S K H C S S S P E F S L D L T R L N L L K R S R I P A H V R P L L

AGTTCTAGAGCGGCGCCACCGCGGTGGAGCTCCAATTCGCCCTATAGTGAGTCGATTACGCGGCTCACTGGCGGCTGTTTACAACGTCGTGACTGG
TCAAGATCTCGCGGCGGTGGCGCCACCTCGAGGTTAAGCGGGATACACTCAGCATAAATGCGGCGAGTGACCGGACGAAATGTTGCAGCACTGACC 7700

→
V L E R P P P R V S S N S P Y S E S Y A R S L A V V L O R R D V

GAAACCCCTGGCGTTACCAACTTAATCGCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCAAGAGGCCCGACCGATCGCCCTTCCCAAC
CTTTTGGGACCGCAATGGGTTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTGACCGCATTATCGCTTCTCCGGGCTGGCTAGCGGGAAGGGTTG 7800

E N P G V T Q L N R L A A H P P F A S V R N S E E A R T D R P S Q

AGTTGCGAGCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGGGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAG
TCAACGCGTCGGACTTACCGCTTACCTGCGCGGACATCGCCGCGTAATTCGCGCGCCACACCAACCAATGCGCGTCGCACTGGCGATGTGAACGGTC 7900

Q L R S L N G E V D A P C S G A L S A A G V V V T R S V T A T L A S

CGCCCTAGCGCCGCTCTTTCGCTTCTTCCCTTCCCTTTCGCGCAGTTTCGCCGGCTTTCGCCGTAAGCTCTAAATCGGGGCTCCCTTTAGGGTTC
GCGGGATCGCGGCGAGGAAAGCGAAAGGAAGGAAGGAGCGGTGCAAGCGGCCGAAAGGGGAGTTTCGAGATTAGCCCCGAGGGAATCCCAAG 8000

A L A P A P F A F F P S F L A T F A G F P R Q A L N R G L P L G F

CGATTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGAATTAGGGTGATGGTTACGTAAGTGGCCATCGCCCTGATAGACGGTTTTTGGCCCTTTGA
GCTAAATCAGGAAATGCGGTGGAGCTGGGGTTTTTGAACCTAATCCCACTACCAAGTGATCACCCTGAGCGGACTATCTGCCAAAAAGCGGAAACT 8100

R F S A L R H L D P K K L D . G D G S R S G P S P . . T V F R P L

CGTTGGAGTCCAGTTCTTAATAGTGGACTCTTGTTCAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGC
GCAACTCAGGTGCAAGAAATFATCACCTGAGAACAAGGTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAAACCTAAATATCCCTAAACGG 8200

T L E S T F F N S G L L F O T G T T L N P I S V Y S F D L . G I L P

GATTTCGGCTATTGGTTAAAAATGAGCTGATTTAACAATAATTAACGGAATTTTAACAATAATTAACGCTTACAATTTAG
CTAAAGCGGATAACCAATTTTTACTCGACTAAATTTTAAATTCGCTTAAATTTGTTTATAATTGCGAATGTTAAATC 8285

I S A Y V L K N E L I . Q K F N A N F N K I L T L T I .

Tuesday, 18 November 1997 13:57

Page 1

fig 55 pCB251 (1 > 8197) Site and Sequence

Enzymes : All 146 enzymes (No Filter)

Settings: Linear, Certain Sites Only, Standard Genetic Code

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCTTASCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA 100
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V

GGAGGTCGCTGAGTAGTGCGCGAGCAAAATTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTAGCGTTTTGCG
CCTCCAGCGACTCATCAGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACGGCTGTTAACGTACTTCTTAGACGAATCCCAATCCGCAAAACGC 200
G G R . V V R E Q N L S Y N K A R L D R O L H E E S A . G . A F C

CTGCTTCGCGATGTACGGGCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGGCTATATGCGCAACTGTAACATAAATGATCAATAATTATCATTAGTTAATGCCCGAGTAATCAAGTATCGGGTATAT 300
A A S R C T G Q I Y A L T L I I D . L L I V I N Y G V I S S . P I Y

TGGAGTTCGCGGTACATAACTTACGGTAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACC TCAAGGCGCAATGATTGAATGCCATTACCGGGCGGACCGACTGGCGGGTGTGTTGGGGCGGGTAAC TGCAGTTATTACTGCATACAAGGGTATCA 400
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGGCC
TTGGGGTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAATGCCATTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGG 500
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A

CCTATTGACGTCAATGACGGTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTGATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGT 600
P Y . R O . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H

TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGCGGTGGATAGCGGTTGACTCACGGGGATTTCGAAGTCTCCACCCATTGACGTCAA
AGCGAATAATGGTACCACCTACGCCAAAACCGTCATGTAGTTACCCGCACTATGCCAAACTGAGTGCCCCCTAAAGGTTGAGAGGTGGGGTAATGCGAGTT 700
R Y Y H G D A V L A V H O W A V I A V . L T G I S K S P P H . R Q

TGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTGACGCAATGGGCGGTAGGCGGTGACGGTGGGAG
ACCTCAAAACAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATGTTGAGGGCGGGTAAC TGC GTT TACCGGCCATCCGCACATGCCACCTC 800
V E F V L A P K S T G L S K M S . O L R P I D A N G R . A C T V G

GTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTGACCGGATCG 900
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A

GTTTAAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAATGGGTCGGGATCTGTACGAC
CAAATTTGAATTGCAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACCTGTCGTTTACCCAGCCCTAGACATGCTG 1000
F K L K L T M G G S H H H H H G M A S M T G G Q O M G R D L Y D

GATGACGATAAGGTACCCGGATCCTTCCGAGACCCACGACGATGTTACGGCTCAGTGCTGCTCCCTGGCTCCAGTGCTCTCCACCTACTCTCAG
CTACTGCTATTCCATGGGCTAGGAAGGCTCTGGGGTGCCTGCTACAAGTGCCGAGTCACGACAGGGACGGAGGTACGGAGGAGGTGGAAGAGGAGTCT 1100
D D D K V P G S F R D P T D D V H G S V L S L A S S A S S T V S S

T7 promoter priming site

Probond binding domain

pCB251 insert = 02

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 2

CTGAGGAGAGGATGCAATCTGAGCAATCCGGAAGCTTCGTAGGGAAC TGGAAATCATCCAGGAAAAAGTGGCCACCTTGACGCTCAGCTTTCTGCCAA
GACTCCTCTCCTACGTTAGACTCGTTTAGGCCTTCGAAGCATCCCTTGACCTTAGTAGGGTCCTTTTTCACCGGTGGAAC TGCAGAGTCGAAAGACGGT 120

pCB251 insert = U2

U2 ORF

A E E R M Q S E Q I R K L R R E L E S S Q E K V A T L T S O L S A H

TGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCTTGGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTSAGCTGCTG
ACGATTAGACCACCGACGAAAAC TCGTCTCGGACCAC TTACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCTGTGACTCGACGAC 130

pCB251 insert = U2

U2 ORF

A N L V A A F E Q S L V N M T S R L R H L A E T A E E K D T E L L

GATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCCAGGCAGTCATTAGGGAGCCCTTAATGCCTCAGAAACACACCCAAAGAAC
CTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTCTTGAGACTCCGGGTCGTCAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTG 140

pCB251 insert = U2

U2 ORF

D L R E T I D F L K K K N S E A Q A V I O G A L N A S E T T P K E

TTCGGATCAAGAGACAAAACCTCTCAGATAGCATCTCAAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAA
AAGCCTAGTTCTCTGTTTTGAGGAGTCTATCGTAGAGTTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCTACGACTACGCTTTTT 150

pCB251 insert = U2

U2 ORF

L R I K R O N S S O S I S S L N S I T S H S S I G S S K D A D A K I

GAAGAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCGGATATA
CTTCTTTTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTCCTCCGGGTTTCAGTCGAAGGAGTATGAGCCTATA 160

pCB251 insert = U2

U2 ORF

K K K K S V V Y E L R S S F N K A F S I K K G P K S A S S Y S D I

GAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAACTACAGCATGGTTCTACAGAGACTGCTTACCCTCCATCAAGTCTCCACCTTGT
CTCCTCAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGTTTGATGTCGTACCAAGATGTCCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAAACA 170

pCB251 insert = U2

U2 ORF

E E I A T P D S S A P S S P K L Q H G S T E T A S P S I K S S T L

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 3

CCTCCGTGGGCACTGATGTCACCGAGGGCCCTGCTCACCAGCCCCACACTAGGCTGTTCCATGCAAATGAGGAGGAGGCCAGAGAAAGAGGCT 120
GGAGGCACCCGTGACTACAGTGGCTCCCGGGACGAGTGGGTCGGGGGTGTGATCCGACAAGGTACGTTTACTCCTCCTCCTCGGTCTCTTCTCTCTCCA
pCB251 insert = U2
U2 ORF
S S V G T D V T E G P A H P A P H T R L F H A N E E E E P E K K E V
ATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAG 180
TAGCCTCGACGCGAGACTCGATACCCCTTCTCTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTC
pCB251 insert = U2
U2 ORF
S E L R S E L V E K E M K L T D I R L E A L N S A H Q L D Q L R E
ACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGG 200
TGGTACGTGTTGTACGTCAACCTCCACC TGGACGACTTTCGTCTCTTACTGGCTGACTTCCATCGGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCC
pCB251 insert = U2
U2 ORF
T M H N M O L E V D L L K A E N D R L K V A P G P S S G S T P G Q
TCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCTGGCACTACCCATTCTTCGGCCCCAGTCTTGACAGACACAGACCTGTACCCAT 210
AGGGACCTAGTAGACGTAATAGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGGTAAAGGAAGCCGGGGTCAGAACGCTGTGTC TGGACAGTGGGTG
pCB251 insert = U2
U2 ORF
V P G S S A L S S P R R S L G L A L T H S F G P S L A D T D L S P H
GGATGGCATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAG 220
CCTACCGTAGTCATGAACACCAGGTTTCTCTTCTACTGGGAGGCCACCACCCTCTACGGGGGGTCTGTGTAGTAGTTTCCCTGAACTTCGTCTGT
pCB251 insert = U2
U2 ORF
D G I S T C G P K E E V T L R V V V R M P P Q H I I K G D L K Q Q
GAATTCTTCTGGGCTGTAGCAAGGTCAAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTCAAGGACTATATTCTAAATGG 230
CTTAAGAAGGACCCGACATCGTTCAGTCACCTTTCAACTGACCTTC TACGACCTACTTCGACAAAAGGTTCAACAAGTTCC TGATATAAAGATTTTACC
pCB251 insert = U2
U2 ORF
E F F L G C S K V S G K V D V K M L D E A V F O V F K D Y I S K H

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 4

ACCCAGCCTCTACCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTGGATGCAGAGCCCCGAGATGCCTCC
TGGGTGGGAGATGGGACCTGATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGCACTTTGTCTACAACCTACGTCTCGGGGGGCTCTACGGAGG 210

pCB251 insert = U2

U2 ORF

D P A S T L G L S T E S I H G Y S I S H V K R V L D A E P P E M P F

TTGCCGTCGAGGTGTCATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATG
AACGGCAGCTCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTACGCAGCTGTCGGACCACAAGCTCTCGGACTAGGGGTTCGGCTAC 220

pCB251 insert = U2

U2 ORF

C R R G V N N I S V S L K G L K E K C V D S L V F E T L I P K P H

ATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGT
TACGTCGTGATGATTTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCGGGGTCGCCGTGCCCGTCTGGATGGACTGGTTAGCGAACC GGCTCA 230

pCB251 insert = U2

U2 ORF

M Q H Y I S L L L K H R R L V L S G P S G T G K T Y L T N R L A E

ACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAA
TGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTTAGACGTTGACATAGAAAGGTT 240

pCB251 insert = U2

U2 ORF

Y L V E R S G R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N

CCTAGCCAACCAAGATAGACCGGGAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTC
GGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCGAGGTAGTCACTCAACCAAG 250

pCB251 insert = U2

U2 ORF

L A N Q I D R E T G I G D V P L V I L L D D L S E A G S I S E L V

AATGGGGCCCTCACCTGCAAGTATCATAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCAACCATGGCTTGCACTTGAGCT
TTACCCCGGGAGTGGACGTTTATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTGGTACCGAACGTGAACTCGA 260

pCB251 insert = U2

U2 ORF

N G A L T C K Y H K C P Y I I G T T N Q P V K M T P N H G L H L S

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1>8197) Site and Sequence

Page 5

TCAGGATGTTGACCTTC TCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTGCTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGC
AGTCCTACAACCTGGAAGAGGTGTTGCACCTCGGTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACG

pCB251 insert = U2

U2 ORF

F R M L T F S N N V E P A N G F L V R Y L R R K L V E S D S D I N A

CAACAAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATC
GTTGTTCTCTCGACGAAGCCCACGAGCTGACCCATGGGTTTCGACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAG

pCB251 insert = U2

U2 ORF

N K E E L L R V L D V V P K L V Y H L H T F L E K H S T S O F L I

GGCCCTTGCTTCTTCTGTGTCGTCCTTGGCATTGAGGACTTCGGGACCTGGTTCATTGACCTGTGGAACAACCTATCATTCCCTATCTACAGGAAG
CCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCC TGGACCAAGTAAC TGGACACCTTGTGAGATAGTAAGGGATAGATGTCCTTC

pCB251 insert = U2

U2 ORF

G P C F F L S C P I G I E D F R T V F I D L V N N S I I P Y L Q E

GAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGTCCGGGACACACTTCCTTGGCCATCAGCCCAACAAGA
CTCGGTTCTACCTATTTCAGGTACCTGTCTTTCGACGAACCTCTCGGTCACCTTACCCAGGCCCTGTGTAAGGGACCGGTAGTCGGGTGTTCT

pCB251 insert = U2

U2 ORF

G A K D G I K V H G Q K A A V E D P V E V V R D T L P V P S A Q Q D

CCAATCAAGCTGTACACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCT
GTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCTAACGGAGTGGAGGGCTCTATCTGTCTAGTTTCGTCTGGGGTTCAAGA

pCB251 insert = U2

U2 ORF

O S K L Y H L P P P T V G P H S I A S P P E D R T V K D S T P S S

CTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCTGGACCCCAACCTTC
GACCTGAGTCTAGGAGACTACCGGTACGACGACTTGAAGTTCTTCGACGGTTGATGTAACTCAGAGGTCTAGTCTTTGGTAGGACCTGGGGTTGGAAG

pCB251 insert = U2

U2 ORF

L D S D P L M A M L L K L Q E A A N Y I E S P O R E T I L D P N L

Tuesday, 18 November 1997 13:57
lig 55 pCB251 (1 > 8197) Site and Sequence

Page 6

[illegible]

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 7

GAATGTGTGTCAGTTAGGGTGTGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGA
CTTACACACAGTCAATCCCACACCTTTCAGGGGTCCGAGGGGTCCGTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTT
N V C Q L G C G K S P G S P G R Q K Y A K H A S Q L V S N O V W I
GTCCTCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAATCCGCCCATCCGCCCCTAAT
CAGGGGTCCGAGGGGTGCTCCGTCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGATTGA
V P R L P S R Q K Y A K H A S Q L V S N H S P A P N S A H P A P N
CCGCCCAGTTCGCCCATTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAAGGCGGAGGCGCCTCTGCCTTGAGCTATTCCAGAAGTAGT
GGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCA
S A Q F R P F S A P W L T N F F Y L C R G R G R L C L . A I P E V V
GAGGAGGCTTTTTGAGGCGCTAGGCTTTTGCAAAAAGCTCCCGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCG
CTCCTCCGAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGAACATATAGTAAAGCCTAGACTAGTTCTCTGTCTTACTCTAGCAAAGC
R R L F W R P R L L Q K A P G S L Y I H F R I . S R D R M R I V S
CATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCCGGCTGCTCTGAT
GTACTAATCTGTTCTACCTAACGTGCGTCCAAGAGGCGGGCGAACCCACCTCTCCGATAAGCCGATCTGACCCGTTGTCTGTTAGCCGACGAGACTA
H D . T R V I A R R F S G R L G G E A I R L . L G T T D N R L L
GCCCGCGTGTCCGGCTGTGACGCGAGGGGCGCCCGTCTTTTTTGCAAGACCGACCTGTCCGGTGGCCCTGAATGAAGTGCAGGACGAGGCAGCGCGGC
CGGCGGCACAAGGCCGACAGTCCGCTCCCGCGGGCCCAAGAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGACGTCCTGCTCCGTCCGCGCCG
C R R V P A V S A G A P G S F C O D R P V R C P E . T A G R G S A A
TATCGTGGCTGGCCACGACGGGCGTTCTTGGCAGCTGTGCTCGACGTTGTACTGAAGCGGAAGGAGCTGGCTGCTATTGGGCGAAGTCCGGGGCA
ATAGCACCGACCGGTGCTGCCCCGAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCCTGACCGACGATAACCCGCTTACGGCCCGCT
I V A G H D G R S L R S C A R R C H . S G K G L A A I G R S A G A
GGATCTCCTGTCATCTACCTTGTCTCTCGCGAGAAAGTATCCATCATGGCTGATGCAATCGCGCGGCTGCATACGCTTGATCCGGCTACCTGCCATT
CCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTCATAGGTAGTACCGACTACGTTACGCCCGGACGATGCGAACTAGGCCGATGGACGGGTAA
G S P V I S P C S C R E S I H H G . C N A A A A Y A . S G Y L P I
GACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGC
CTGGTGGTTTCGCTTTGTAGCGTAGCTCGCTCGTGCATGAGCTTACCTCGGCCAGAACAGCTAGTCTTACTAGACCTGCTTCTCGTAGTCCCCGAGCGG
R P P S E T S H R A S T Y S D G S R S C R S G . S G R R A S G A R A
CAGCCGAATGTTGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAA
GTCGGCTGACAAGCGGTCCGAGTTCCGCGCGTACGGGCTGCCGCTCTAGAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACACCTTTT
S R T V R O A Q G A H A R R R G S R R D P V R C L L A E Y H G G I
TGGCCGCTTTTCTGGATTATCAGCTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGC
ACCGGCGAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCTGGCGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACCG
W P L F W I H R L V P A G C G G P L S G H S V G Y P . Y C . R A W
GGCGAATGGGTGACCGCTTCTCTGCTTTTACGGTATCGCGCTCCCGATTGCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTTCTTCTGAGCGG
CCGCTTACCCGACTGGCGAAGGAGCAGAAATGCCATAGCGGCGAGGGCTAAGCGTCCGCTAGCGGAAGATAGCGGAAGAACGCTCAAGAGACTCGGC
R R M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1>8197) Site and Sequence

Page 8

GACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCAGAGATTTCGATTCCACCGCCGCCCTCTATGAAAGGTTGGGCTTCGGAATC
CTGAGACCCCAAGCTTTACTGGCTGGTTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACCTTCCAACCCGAAGCCTTATG 580X
T L G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I
GTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACA 590X
CAAAAGGCCCTGCGGCCGACCTACTAGGAGGTGCGGCCCTAGAGTAGCAGCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTGGAATATTACCAATGT
V F R D A G V M I L Q R G D L M L E F F A H P N L F I A A Y N G Y
AATAAGCAATAGCATCACAAATTCACAAATAAGCATTTTTTCTACGCTTCTAGTTGGGTTGTCCAAACTCATCAATGTATCTTATCATGTCTG 600X
TTATTTCTGTTATCGTAGTGTTTAAAGTGTTTATTTCTGAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATAGTACAGAC
K . S N S I T N F T N K A F F S L H S S C G L S K L I N V S Y H V C
TATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATGTTATCCGCTCACAATTCACACAACATACGAGCC 610X
ATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACCAGTATCGACAAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTTGATGCTCGG
I P S T S S . S L A . S V S . L F P V . N C Y P L T I P H N I R A
GGAAGCATAAAGTGAAAGCTGGGGTGCTAATGAGTGAGCTAATCATTAATTCGCTTGGCGTCACTGCCCGCTTTCAGTCGGGAAACCTGTGCT 620X
CCTTCGTATTTACATTTTCGACCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCA
G S I K C K A V G A . . V S . L T L I A L R S L P A F Q S G N L S
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGGCTATTGGGCGCTCTTCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGT 630X
CGGTCGACGTAATTACTTAGCCGGTTGGCGGCCCTCTCCGCCAACGCATAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCA
C Q L H . . I G O R A G R G G L R I G R S S A S S L T O S L R S V V
TCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCSTGAATACGGTTATCCACAGAATCAGGGGATAACGAGGAAAGAACATGTAGCAAAAGGCCAG 640X
AGCCGACGCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATAGGTGCTTAGTCCCTATTGCGTCTTTCTGTACACTCGTTTCCGGTC
R L R R A V S A H S K A V I R L S T E S G D N A G K N M . A K G Q
CAAAAGGCCAGGAACCGTAAAGGCCGCGTTGCTGSCGTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGST 650X
GTTTTCGGTCTTGGCATTTCGCGGCCAACGACCGCAAAAGGTATCCGAGGCGGGGGAGTGCTCGTAGTGTTTTAGTCGCGAGTTACGCTTCCA
Q K A R N R K K A A L L A F F H R L R P P D E H H K N R R S S Q R
GGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCGCGCGCTTACCGGATACCTGT 660X
CCGCTTTGGGCTGTCTGATTTCTATGGTCCGCAAGGGGGGACCTTCGAGGGAGCACGCGAGAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAG
V R N P T G L . R Y Q A F P P G S S L V R S P V P T L P L T G Y L S
CGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAGCTGTAGGTATCTCAGTTTCGGGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCA 670X
GCGGAAAGAGGGAAGCCCTTCGCACCGGAAAGAGTTACGAGTGCGACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTTCGACCCGACACAGTG
A F L P S G S V A L S O C S R C R Y L S S V . V V R S K L G C V H
GAACCCCGGTTAGCCCGACCGCTGCGCCTTATCCSGTAACATATGCTTGTAGTCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG 680X
CTTGGGGGCAAGTGGGCTGGCGACGCGGAATAGGCCATTGATAGCAGAACTCAGGTTGGGCCATTCTGTGCTGAATAGCGGTGACCGTCTGCGTGAC
E P P V Q P D R C A L S G N Y R L E S N P V R H D L S P L A A A T
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTG 690X
CATTTGCTCTAATCGTCTCGCTCCATACATCCGCCACGATGCTCAAGAATTCACCACCGGATTGATGCCGATGTGATCTTCTGTCTAATACCATAGAC
G N R I S R A R Y V G G A T E F L K V V P N Y G Y T R R T V F S I C

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 9

CGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTCAGGCA3
GCGAGACGACTTCGGTCAATGGAAGCCTTTTTCTCAACCATCGAGAAC TAGGCCGTTTGTGGTGGCGACCATCGCCACCAAAAAACAAACGTTCTGTC
A L L K P V T F G K R V G S S . S G K Q T T A G S G G F F V C K Q 750

CAGATTACGCGCAGAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTTGG
GTCTAATGCGGCTTTTTTCTAGAGTTCTCTAGGAACTAGAAAAAGTGGCCAGACTGCGAGTCACCTTGCTTTTGAGTGCAATCCCTAAAAACC
Q I T R R K K G S Q E D P L I F S T G S D A Q W N E N S R . G I L 760

TCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAACTTGGCTGACAG
AGTACTCTAATAGTTTTTCTAGAGTGGATCTAGGAAATTTAATTTTACTTCAAAATTTAGTTAGATTTATATATCTCATTGAACAGACTGTCT
V M R L S K R I F T . I L L N . K . S F K S I . S I Y E . T W S D S 770

TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCCATAGTTGCCTGACTCCCGCTCGTGTAGATAACTACGATACGG
AATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACATCTATTGATGCTATGCC
Y O C L I S E A P I S A I C L F R S S I V A . L P V V . I T T I R 780

GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCG
CTCCCGAATGGTAGACCGGGTACGACGTTACTATGGCGCTCTGGGTGCGAGTGCGCGAGGTCATAATAGTCGTTATTTGGTCCGGTCCGCTTCCCGG
E G L P S G P S A A M I P R D P R S P A P D L S A I N O P A G R A 790

AGCGCAGAAGTGGTCC TGAACCTTTATCCGCCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTGCGCAA
TCGCGTCTTACCAGGACGTTGAAATAGGCGGAGGTAGGTGAGATAATTAACAACGGCCCTTCGATCTCATTATCAAGCGGTCAATTATCAACGCGTT
E R R S G P A T L S A S I Q S I N C C R E A R V S S S P V N S L R N 800

CGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGGCTTCAATCAGCTCCGGTCCCAACGATCAAGGCGAGTTACATGATCC
GCAACAACGGTAACGATGTCGTAGCACCACAGTGCAGCAGCAACCATACCGAAGTAAGTCGAGGCCAAGGGTGTCTAGTTCGCTCAATGACTAGG
V V A I A T G I V V S R S S F G M A S F S S G S Q R S R R V T . S 810

CCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCGAGTTATCACTCATGGTTATGGCAGCACTGC
GGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACAGTCTTCAATCAACGGCGTCACAATAGTGAGTACCAATACCGTCTGTGACG
P M L C K K A V S S F G P P I V V R S K L A A V L S L M V M A A L 820

ATAATTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTG
TATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCAC TCATGAGTTGGTTCAGTAAGACTCTTATCATATACGCCGCTGGGCTCAAC
H N S L T V M P S V R C F S V T G E Y S T K S F . E . C M R R P S C 830

CTCTTGGCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGCTCATCTTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGS
GAGAACGGGCGCAGTTATGCCCTATTATGGCGGGGTATCGTCTTGAATTTTACGAGTAGTAACCTTTTGCAAGAAGCCCGCTTTTGAGAGTTCTC
S C P A S I R D N T A P H S R T L K V L I I G K R S S G R K L S R 840

ATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAA
TAGAATGGCGACAACCTTAGGTCAAGCTACATTGGGTGAGCAGTGGGTGAC TAGAAGTCGTAGAAAATGAAAGTGGTCGCAAGACCCACTCGTTTTT
I L P L L R S S S M . P T R A P N . S S A S F T F T S V S G . A K 850

CAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCAATTATCAGGS
GTCTTCCGTTTACGGCGTTTTTCCCTTATTCCCGCTTGCCCTTACAACCTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAATAGTCCC
T G R O N A A K K G I R A T R K C . I L I L F L F O Y Y . S I Y O G 860

Tuesday, 18 November 1997 13:57
fig 55 pCB251 (1 > 8197) Site and Sequence

Page 10

. TTATTGTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGACGTC
AATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTATTGTTTATCCCAAGGCGCGTGTAAGGGGCTTTTCACGGTGGACTGCAG 8197
Y C L M S G Y I F E C I . K N K Q I G V P R T F P R K V P P D V

Figure 56

Monday, 1 December 1997 14:13
fig 56 pNP10

Page 1

10 20 30 40 50 60 70
AAGCTTGCATGCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCCACCCTGAGGATCAGAAGAAAT 70
TGGAGCAACTACCCACATCCATTATGCCACCCGCGGTTCTAAGTGAGTTTAATTTGAGTTTACGACGA 140
CAAAAATGTGTTCCTTAAAGAACATCTTCGACCTGAGCTATCTGTATGACTAGTTGTTGAGTGATTTT 210
TCATGAGAGAAATATTAAGGAACATTATTTACTTTGCTTATTTGCCCTAACTTTGATTTAGTTTTTCC 280
ATCAACTAGATCTTACAAAACCTTGCATACAAATTCATTTTCAGATTACCGTCCGACGTTGCGCCACGT 350
360 370 380 390 400 410 420
CAGCAACCGCTTCAGCAACTAACCCAAATTCGAATTTCCACAAATGTCAACATCCAGGCCTCAGACTCC 420
ACAGTCAAGAAATATCGAAAATTTGGTAAGAAATTTATTTTGAGCTCAAACCTGTATATAAATGCCAGAAA 490
GAAGATGATAAAAAATGAGTTTTCGAAAACCTTCCACCTTTATTTGCTCTAATAAGACGGCTTATATCT 560
CAATTTTCTTGAGTTTATCAAAAAATTTCCACTATACAAATGTAGAAAAGTATTTTGCACAAATTTTG 630
TCAGTTGACAGCTTTGTAATAGATCCAAATGGAACTTAGATACAAGCTGTAAAGTGGAAAGGACCGCAAG 700
710 720 730 740 750 760 770
TCATATCTGGAAATATGATCTGAAACAAATTTGTGCTATTCTCAAAATGTTTAAAGCATGTTTGAAGAT 770
TTTTTCAAATTCGACATGTTTCAGAACCTTCTTTTTTGTATGAAAAAGTAAAAAATACTATTTCAAAC 840
CCTCACCGCCACCAATTTCAACTCTTAATTTTATAAAATTTTGCAATTTACAATTCGCTTCCCTTGC 910
CCGAAAAGTGCCCAACCAAAATCAATTTCTGGCTTCATAATGACTTTAAATTGATGTGAGAAAACACAG 980
AAGAGGCTAACTAAATTCAGACAGGACAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050
1060 1070 1080 1090 1100 1110 1120
TCCATCTCCAACAACAACAATTTTCCAAATTCGCTTGTCCATTTTGTCTTATAAACAATTTGTGTGTGCAAGG 1120
AAACTACACCGGAGAGCGGTCAATTAATTCGAATGAGAGCATGGCAATTACCTCTTCGGAATTTGATGAA 1190
TAAAGATAGAGCGGATGACACTGGCTGCTAGTAGTATGAGTGTAGAAATTTGCTTTTTCATGCTCTCAAC 1260
CGGCAAGAGTCTTCCCT 1330
CCACATCTCTGGCTTACCACAAACTGGAATACATTTTACACTATTCAGCCATTTATTTTGAATTTTA 1400
1410 1420 1430 1440 1450 1460 1470
ATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAGTTTATTTAAAAAACGATATTTTCGATTTTAAA 1470
AAATCTGAAAAGTTTCAAAAAATCAATTAATTTCCCTAACAAATTTGATGGCTAAATTTTATTTCTAC 1540
TGTTCACAAATATCTTTATATGTATCACTGTTTTCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAAG 1610
CTCCGTCTCACATTTCCCAATGCTATGATCGCTACTCAGCACATATCCAAAAATTAAGCTAGACGCTTGA 1680
TAAATATTGGGCACGCGTAATAAAGTGAAGCAGTTAGAATTTTAATCAAGCACAGATTATCTATCAAA 1750
1760 1770 1780 1790 1800 1810 1820
TTCAAATCTTTGAACATTCAGCCAGTTCTGACAAATTTCCATGCTTTTTCGCCCATTAATAAAGCTTTCTCA 1820
CTCTCTCATCCATCTCACTCGTATCATAAAAGTATAGCAAAAGCCCGACCTGACCTTTTAAAGAGAAAGGA 1890
GATACTGAGCCACAATGGCGTGTGACCTTTTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1960
CTCTTCAAAATAGCCATAGACCTCTCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2030
TGAAGCCCGGGAATAATAGTATATTTAGAGCTTATCTTTATGCAATACATAAAATAAGAGGCAATTTTA 2100
2110 2120 2130 2140 2150 2160 2170
AAATAATTAATAATTAATGAGGTTTGTAGATGTAGATTTCGAAAAGAAGAAAAAACAATAACAAATAGGAAC 2170
CGCCAGATCAAAATCTTAAAGCTTTTCAAGATTTTAAAGCAAGATTCGGCTCAACAGAAAACTGAA 2240
GTCCCTGCATAAAATCAAGTGAACGTTTAGATTGAACCTCGAAATCCTAAGCCTGACATATAGCTTAACT 2310
CTAGATCTTAGTTGCCCAATAGCTCAAGCCCAAGCAGAAAAGACTTGCATTTAGTTTAAAGCCTAGATTGA 2380
CTTGCCTGCTTCACTCAATCCAGACATGATTTCCAAAGAGAGTTTCAATTTTAATTTTCCAGTTTCT 2450

Figure 5b

Monday, 1 December 1997 14:13
fig 5b pNP10

Page 2

2460 2470 2480 2490 2500 2510 2520
TGTTACTTAAATCTTAATGCCGCGTGATGCGTAAATCGTTATCCCTTTCTCTCACACTTTCATTTACA 2520
GATTCATCAAAAGATGGTATCAAGCCAAAGACGTCTGGACGAAACCACCCCATCATCAACCACGCAI 2590
CAAAATATACAAATTCATTCCGTCCGTCCGAGCCGTCGAGTGGCAATAATAATGTTGGCTCCACCATATC 2660
CACATCTGCCAAGAGCTTAGGTATCCGATCCTTCGGGCTTCTTTTAGAAATGAGAGGAGGAGGAGGAGG 2730
TCATCAACGTACAGCTCTATTTGGAAGCTAAACCGACCTACCTCCCAACTCCAAAACCTTCTAGACCAC 2800
2810 2820 2830 2840 2850 2860 2870
AAACCCAGCTAGTTCGTGTTGCTACAACCTACAAAAATCGGAAGCGCAAAGCTAGAGGATCCCCGGGATGG 2870
GCCAAAGGACCCAAAGGATGATGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2940
AGGGTACCGGTAGAAAAATGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3010
TAGATGGTGTATGTTAATGGGCACAAATTTCTGTCACTGGAGAGGGGGAAGGGAAGGGAAGGGAAGGGAAG 3080
ACTTACCCTTAAATTTATTTGCACTACTGGAAAACACCTGTTCCATGGGTAAAGTTAAACATATATATA 3150
3160 3170 3180 3190 3200 3210 3220
CTAACTAACCCCTGATTATTTAAATTTTCAAGCCAACTTGTCACTACTTTCTGTTATGGTGTTCAATGCT 3220
CTCGAGATACCCAGATGATGAAACCGGCATGACCTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA 3290
GGAAAGACATATATTTTCAAGATGACGGGAACCTACAAGACAGGAAGGTTAAACAGTTCGGTACGAAC 3360
TAACCAACATATGAAATGCAAGTGCAGAGTCAAGTTGAAGGTGATACCTTGTTAATAGAATCG 3430
AGTTAAAGGATGATTTAAAGAAGATGAAACATCTTGACACAAATTGGAAATACAACTATAACTC 3500
3510 3520 3530 3540 3550 3560 3570
ACACAATGTATACATCATGCGAGACAAACAAAAGAATGGAAACAAAGGTTGGAAGTTGAAACCTGGACTGA 3570
CTAACGAACGGATGATGAAAGTTGCAAGACTTCAAAATTAGACACAACATTGAAGATGGGAAGCGTTC 3640
AACTAGCAGACCATATCAACAAAATCTCCAAATGGCGATGGCCCTGTCTTTTACCAGACAAACCATTA 3710
CCTGTCCACACAATCTGCCCTTTCGAAAGAATCCCAACGAAAGAGAGACACATGGTCTTCTTGAGTTC 3780
GTAACAGCTGCTGGGATTACACATGGCATGATGAACGATACAAATAGCATTTCGTAGAATTCCAACCTGAG 3850
3860 3870 3880 3890 3900 3910 3920
GGCGGGTCGCTACCAATACCAACCTGTCTGGTGTCAAAAATAATAGGGGCGCTGTCACTCAGAGTAAGTT 3920
TAAACTGAGTTCTACTAACTAACGAGTAATATTAATTTTCAAGCATCTCGCGCCCGTGGCTCTGAGCTTC 3990
TAAGTCCAAATGAGTCTGCAAGATCGGTACATGGTCTTTCTGCTGTGGTGGGAGGAGGAGGAGGAGGAGG 4060
TATCAAAAAAACTTCTTTAATTTCTTTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4130
TGTAGATTCAAAAAAGAAATTAATTCGTAATAAAAAAGTCGAAAAAAATTTGTGCTCCCTCCCCCATTAAT 4200
4210 4220 4230 4240 4250 4260 4270
AATAATTTCTATCCCAAAATCTACACAATGTTCTGTGTACACTTCTTATGTTTTTCTTCTTCTTCTTCT 4270
TTTTTTTGAACATCATAGAAAAAACCCACACAAAAACCTTATCATATGTTACGTTTCAAGTTTATGAG 4340
CGCAATTTTTATTTCTTCCGACGTCGGGCTCTCATGACGTCAAAATCAATGCTCAATGCTGAAAAAGTTT 4410
GGAGTATTTTTGAAATTTTCAATCAAGTGAAGTTTATGAAATTAATTTCTGCTTTTCTTCTTCTTCTTCT 4480
GTTTTCCCTATTTGTTCTCAAGAGTTTCTGAGGACGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4550
4560 4570 4580 4590 4600 4610 4620
GCACGATGCAAGAAATCTCGGAAGAAGGTTTGGGTTTGAGGCTCAGTGGAAGGAGGAGGAGGAGGAGGAGG 4620
TTTGAAAGTGGAGTAGTGTCTATGGGGTTTGGCTTAAATGACAGAAACATGCCCAATATACCAAAACA 4690
TAACGTTTTCT 4760
TCTGACACATGCAAGTCTCCGGAGACGGTCAACATTTGTCTGTAAAGCGGATGCCGGGACGAGACAAAGCCG 4830
TCAGGGCGGTCACGGGTTTGGCCGGTGTCTGGGGTGGGTTAACTATGCCGCATCAGAGCAGATTTGTA 4900

Monday, 1 December 1997 14:13

Page 3

lig pNP10

4910 4920 4930 4940 4950 4960 4970
CTGAGAGTGCACCAFAFGCGGFGTSAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGGC 4970
CTTAAGGGCCFGFGAFCGCGCTATTTTATAGGTAAATGTCATGATAATAATGGTTTCTTAGACGTCAG 5040
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTGTTTTATTTTCFAAFAACATTCAAAATATGTA 5110
TCCGCTCATGAGACAAFAACCTGTATAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA 5180
CAITTCGCTGTCGCCCTTAATCCCTTTTTCGGGCATTTTGCCTTCCTGTTTTTGTCTACCCAGAAAGCG 5250
5260 5270 5280 5290 5300 5310 5320
TGGTGAAGTAAAGATGCTGAAGATCAGTGGGTCACGAGTGGGTTACATCGAACFGGAATCTCAACAG 5320
CGGTAAGATCCTTGAGAGTTTTTCGCCCGGAAGAAGCTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA 5390
TGTCGGCGCGGTATTATCCCGTATTGACGCGGGCAAGAGCAACTCGGTGCGCGCATACACTATTCTCAGA 5460
ATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG 5530
CAGTGC TGGCATAACCAFGAGTGAFAACACTGCGGCCAAGTTACTTCTGACAACGATCGGAGGACCGAAG 5600
5610 5620 5630 5640 5650 5660 5670
GAGCTAACCGCTTTTTCGACAAACAAGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA 5670
ATGAAGCCATACCAAAACGACGAGCGTGACACCACGATGCCFGTAGCAA TGGCAACAACGTTGCGCAAACT 5740
ATTAAGTGGCGAAGTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGCGGATAAAGTT 5810
GCAGGACCACTTCTGCGCTCGGCCCTTCGGCTGGCTGGTTTATTGCTGATAAACTCGGAGCGCGGTGAGC 5880
GTGGGTC TCGCGGTATCATTCAGCACTGGGGCCAGATGGTAASCCCTCCCGTATCGTAGTATCTACAC 5950
5960 5970 5980 5990 6000 6010 6020
GAGGGGGAGTCAGGCAACTATGGATGAACGAATAGACAGATCGCTGAGATAGGTGCGCTCACTGATTAAG 6020
CATTGGTAAC TGCAGACCAAGTTCAC TCA TATACTTTAGATTGATTTAAACTTCATTTTAAATTA 6090
AAAGGATC TAGG TGAAGA TCC TTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCA 6160
CTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCCTTTTTTTCGCGCGTAAATCTGC 6230
TGCTTGCAAAACAAAAAAGCACCGGTACCAAGCGGTGGTTTGTGTTGCGCGATCAAGAGCTACCAAC TCTT 6300
6310 6320 6330 6340 6350 6360 6370
TTCGGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAAATAC TGTCTTCTAG TGTAGCGGTAGT TAAAG 6370
CCACCACTTCAAGAACTCTGTAGCACCGGCTACATACC TCGCTC TGC TAA TCC TGT TACCAAGTGGCTGCT 6440
GCCAGTGGCGATAAGTGGTGTCTTACCGGG TGGAC TCAAGACGATAGTTACCGGATAAGGCGCAGCGGT 6510
CGGGCTGAACCGGGGGGTCTGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT 6580
ACAGCGTGAGCATTGAGAAAGGCGCACGCTTCGCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGCG 6650
6660 6670 6680 6690 6700 6710 6720
AGGGTCGGAAACAGGAGAGCGCACGAGGAGCTTCCAGGGGGAACGCC TGG TATCTTATAGTCTGTCTC 6720
GGTTTCGCCACCTCTGACTTCAGCGTCCA TTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAATTA 6790
CGCCAGCAACCGCGGCTTTTACGGTTCTTGGCTTTTTC TGGCTTTTGGCTCAGATGTTCTTTCTCTCG 6860
TTATCCCTGATCTGTGGA TAAACCGTATTACGCGCTT TGAAGTGAAGCTGATACCGCTCGCCGACGCGAA 6930
CTACCGAGCGCAGCGAGTCAATGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAAACCGGCTCTCCCGC 7000
7010 7020 7030 7040 7050 7060 7070
GGCTTCGCGATTCATTAATGCAGCTGGGACGACAGCTTTCCCGAG TGGAAAGCGCGGCACTGACCGCAAC 7070
GCAATTAATGTGAGATTAAGTCACTCATTAGGCACCCAGGCTTTACAC TTTATGCTTCCCGCTCGTATGT 7140
TGTGTAATTGTGAGCGGA TAAACATTTACACAGGAACAGCTATGACCATGAT TACGCCAAGCTGTA 7210
AGTTTAAACATGATCTTAC TAACTATTCTCATTTAAATTTTACAGAGCTTAAATATGGCTGAAATCA 7280
CTCACAACGATGGATACCTAACAAT TGGAAATGAAAT 7319

Figure 57

Monday, 1 December 1997 14:12

ig PCB501

Page /

10 20 30 40 50 60 70

ATACCAAGATACGCAAGCTTGCATGCCTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCT 70
AGAGGATCAGAAAGAAATTCGACCAACTACCCACATCCAATTCGACCCCGGGTTTCGAGTGAGTTAA 140
TTTGTAGTTTACGACACAAAAATGTGTTCTTTAATACTATCTTCGACTTGAGTCATTCTGTATGACT 210
AGTTGTTGAGTGATTTTTCATTGAGAAAAATATTAAGGAACATATTTACTTTTGCATTTTGCCTTAA 280
TTTGATTAGTTTTCGATCAACTAGATCTTACAAAACTTGCAATACAATTCATTTCAGATTACCCCTC 350

360 370 380 390 400 410 420

GCACGTCGCGCACGTCAGCAACCGGTTGAGCAACATACCCCAAATTCGCAACTTTCGACAAAATGCAACA 420
TCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAAATTCGTAAGAAATTTATTTTGGAGCTCAAACTTGT 490
ATAAAAATGCCAGAAAAAGAGATGATAAAAAATGTAGTTTTCGCAAAAACTTCCACCTTTATTGCTCTAA 560
TATGACGGCTTATATCTCAATTTCTTGAATTTATCAAAAAATTTTCGACATACAAAAGTATGAAAAAGT 630
ATTTTCGACAAAATTTTCGATGATGACAGCTTTTGAATAGATCCAAATGGAACCTAGATACAAAGCTGTTAA 700

710 720 730 740 750 760 770

AGTGGAGGAGCGCAAGTCTATACATGGAAAAATATGATCTGAAACAAATTTGTGCTATTCTCAAAATGTTTA 770
AGACATGTTTTGAAGATTTTTCGCAATTCGCACTAGTTTCAGAACCTTCTCTTTTGTATGAAAAAGTAAA 840
TAAAAAATTTTCAAACTTCACCGCCACCATGTTTCGCACTCTTAATTTTATAAAATTTTGCATTTTAC 910
AAATCGCCCTCCCTTTCGCGGAAAAAGTCCCAACCAAAATCAATTTCTCGGCTTCATAATGACATTTTAAAT 980
TATGTCGAAAAACACAGAAAGAGGCTAACTAAATTCGACAGGGGACAGGTTGTCCCTCTTTCGCGCTCTCT 1050

1060 1070 1080 1090 1100 1110 1120

CCGCTCCTCCTCCCTCCCTTCGATCTCCAACACAACAATTTCCGAATTTCTGTTGTCCATTTTCGCTTA 1120
TATTTGTGTTGGAAAGGAAACTACACGGGAGACGGTCAATTAATTCGAAATGAGAGCAATGGCAATTTAC 1190
TTTCGGAAATTTGATGAATAAAGATACACCGGATGACACTGGCTGGTAGTAGTATGAGTGTAGATTGCTT 1260
TTTCATCGTCTCAACTTGGGCAATGAGTCTTTCGCGCTCTCATCACTGACAATTAATGTGCGGGTTTATG 1330
CGCTCTTTCTTATTCGCGCACTCATCTGCGGTTACCAACAACTGGAATACATTTTACACTACTATTTCAAGCC 1400

1410 1420 1430 1440 1450 1460 1470

ATTTATTTTCATATTTAAATTTTGTGCAATTAGGGATAAACACGACTTTTAAAAAGTTTATTTAAAAAAACG 1470
ATATTTTCGATTTTAAAAAAATTCGAAAAATTTCAAAAAATCAATAAATAATTCGCAACCAAAATTTGATGCC 1540
TAAAAATTTTATTTCTACTGTTGACAAATATCTTTATATATGATCACTGTTTTCATCTCAAAACCTTGAAAT 1610
CCCCAAGTTATAGGAAGCTCCGTGTGACATTTCCCATGCTATGAATCGCTACTCAGCACATATCCAAAAA 1680
TTAAGCTAGACGGTTTATAATTTATTTGGSCACGCTAATAAAGTGAAGGCAATTTAGAAATTTAATTCAAGC 1750

1760 1770 1780 1790 1800 1810 1820

ACAGATTATCTATCAAAATTCAAATCTTTGAAACATTCACCCAGTTCTGTACAAATTTTCATGCTTTTGGGCC 1820
ATTAATAAAGCTTTTCGACCTCTTCATCCATCTCACTCGTATCATATAAAGATATAGCAAAAGCCCGACAT 1890
ACTTTTAAAGAGAAAGGAGATACGAGGACCAATGGCGTGTGACCTTTTCATCTCGTCCGTTCGGCTCAA 1960
ATTCAGGCTCAATCAATCTTCAATATGCCATAGACCTCCTTGTCTTCTTCGTTTGTACTCGCGCC 2030
TATTTTTCGTTGCTGCTGAAGGCGGGGAAAAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATA 2100

2110 2120 2130 2140 2150 2160 2170

AAAAACGAGGCAATTTAAAAATTTAAATTAATGAGGTTGTAGATGTAGATTTGGAAAGAGAAAAAA 2170
ACAAAAACAAATAGGAACCGCCAGATCAAAATTTCTATTTAAAGGTTTTCAGATCTTTAGGCAAGATTCGG 2240
CTAAACAGAAAACTGAAGTGCCATGATAAATCTAGTGTAACTTTAGATTGAATCGGAAATCCGAAAGCC 2310
TGAATATAGCTTATTTCTAGATCTTAGTTGCGCAATAGCTCAAGCCCAAGCAGAAATGACATTCATTTA 2380
GTTAAGCCATGATGACTTCTTGTGCTGAGTCTAATCCAGACTAGATTTCCAAGAGATTTTCAATTTT 2450

Monday, 1 December 1997 14:12
flg pCB501

Page 2

2460 2470 2480 2490 2500 2510 2520
AAATGTTTCCAGTTTCCTGGTACCTAAATCTTAATGCCCTGTGATGCGTAAATCGTTATCCCTTTCTC 2520
TCACACCTTCAATACAGATTATCMAAGATTGGTATCAAGCCAAAGACGCTGGAGCTTAAACCACTTC 2590
ATCATCAACCACCTCATCAAAATAACAAATCATTCCTCCGTCGAGCCCTTCGAGTGGCAATAATAAT 2660
GTTGGCTCGACGATATCCACATCTGCGAAGAGCTTAGGTATCCGATCCTTCGGCTTCTTTTATAGAAAT 2730
ATATTATTTTCAGATCATCATCAACGTACAGCTCTATTTCGAATCTAAACCGACCTACCTCCCAACTCCA 2800
2810 2820 2830 2840 2850 2860 2870
AAAACTCTCTAGACCACAAACCCAGCTAGTTCGTGTTGCTACAACACAAAAATCGGAAGCTCAAGGCTA 2870
GCCGCTCCGAAAGCCGTGAGCACCCCAAACTTGCTTCCTGTAAGACATCTGGAGCAAAACAAGAGCCCG 2940
ATAACAGCGGT 3010
ATCGAATAGCCCAACACCTACGAGAAAGGCGGCGGCGGCTGCCTCAACAACAACTTTGTCGAAAAATCGCT 3080
GCCCGAGTGAAGAAGTGGCTGAAGCCCGCCGACCAAGCTGGGAAGTGGCAGCTCTATGTGGAAGCTTT 3150
3160 3170 3180 3190 3200 3210 3220
GTACGCCAAAAGTTTCTTACCTTAAACCGGACGCCCAATCATATCTCAACAAGACTCGAAACGATGCTC 3220
AAAGAGCAGTGAAGAAGAGTCCGGATACGCTGCATTCAACAGCAGCTCGCCAACTCATCATCGACGGAA 3290
GGTTCCCTAAGCAATGCATTCCCATCTTCCAAGAGTTCACAGCTCAGACGAAAAAGTCTCCGTCATCAGACG 3360
ATCTTACTCTTAACGCCCTCCATCGTGACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAA 3430
TATTATCAACAAGCTCTTGAGGAAAAACCAACACCTGGCAGTGAAGGAGTGAAGAAACACAGCGAAAAAA 3500
3510 3520 3530 3540 3550 3560 3570
GATCCACTCCAGCTGTTCGGCCACCTGACACCCAGCCAAACAATCGGAGTTGTTAGTCCAATTATGGCAC 3570
ATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACAGAACTGAAAAAGCTCCAATCAATGAGCAT 3640
CGACACGACGGACGTTCCACCGCTTCCACCTCTCAAAATCACTTGTTCCTTAAAAATGACTTCAATCCGA 3710
CAACCCACCAACCTACATGTTCTTCTAAACAAGGAAAAATCATATCGCCTGTCAAGTCTGTTGGATATG 3780
AGCAGTCTGTCGGCTCTGAAGACTCCATTGTGCTCATGCTCGGCTCAGGTGACTCCGCCGACAAAAAC 3850
3860 3870 3880 3890 3900 3910 3920
TTCTGGTAACTCATTCCTGAGAGAAAGATGAGGAAAGAATAAGACATCAGAATCCAGCGGCTACACCTCT 3920
GACGCGGGTGTGCGATGTGCGCCAAAATGAGGGAGAAAGCTGAAAGAATACGATGACATGACCTCGTGGAG 3990
CACAGAACGGCTATCTTGACAACTTCGAAGACAGTCTCTCTTGTCTCTGGAATATCCGATAAACAACA 4060
GCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGGCAACAGTCGCCGCCAAACAATGCGAC 4130
TATTCCACTTTGTTCGCCATCCGACGCTTCTTCTCTCAAAAGCCCGAGTCCCCAGTCCGGTCTCCACAT 4200
4210 4220 4230 4240 4250 4260 4270
CAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTACAACTTCTGTCTCCAGTGGCCGAACGAGCCA 4270
ACCTGGCGCGCTGCCACCTCAACCTTCGGACAACATTCGCTAAGATCCCGGGGATCTCATCTTATCTCT 4340
CCACACTTATCAGTCTCAGCTGATAAGGACACAATGTCTATCCACTCACAGACTAGTGGACGACCTCTCT 4410
CACAAAAACCAAGCTATTTCAGGCCAATTTCTATCTACTTGTCTGTAATGCCACCTTCAAGAGTTCACATC 4480
CACCGAGCACAGAATGGCGGCTCTCTTGAGCCCGAGACGGGTGCCGAACCTGATGTGGAATATGATTCT 4550
4560 4570 4580 4590 4600 4610 4620
TCAGGATCTTACTCGGCGCTCTCTCCGAGGTAGAAAGCTCTACTGGTATCTTATGGAGAGAGCTTCTAACCTCT 4620
ACAGACTATCCGATGAAGAAATCCCGCGCACATCTCTGCCAAAAGTGAGATGGGAATCCCAACCTGCTACTGCG 4690
TAGCACGACAGCATATGGAATCTCTCAATGAGAAGTACGAACATGCTATTTCGGGACATGGCACGCTACTTG 4760
GAGTGTACAGAAGAACCTGCTGACTCACTAAGCAAGAAACAGGAGAAGCTATGGAGCATGCTGTTGATCTCT 4830
TTGAGCAAAAGCTTAGAAAACTCACTCAACACATGATCGATCCAACTTGAAGCTTGAAGAGGCAATACG 4900

Monday, 1 December 1997 14:12

Page 3

Ilg pCB501

4910 4920 4930 4940 4950 4960 4970
A T CAGGCAGGACATTGCTCATTTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAAC 4970
GAAGGCGCTGGTGAGCTTCTTCGTCAACCA TCTGGAATCAGTTGCATCCATCGATCATCGATGTCAT 5040
CGTGTGCAAAAAGCAGCAAGCAGGAGAAGATCAGCTTGAGCTCGT TGGCAAGAACAAGAAAGAGCTGGAT 5110
CGCTCTCTACTCTCCAAG TACCAAGAAGAAGAACTACGACGAAGCAGATATGCCATCAATT 5180
TCCGATCTCAAGGAAC TCTTGACAACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAAGCGGATA 5250
5260 5270 5280 5290 5300 5310 5320
GTGCACTTTACGAAG TCCGCT TGGCAATCTGGATCGTSCCCGCGAAGTTGATG TCTGAGGGAGACAGT 5320
GAATTAAG TGGAAATCTGAGAACAGCAATTAAAGAAAGAGTGGACAAACTCACCACCGTCCAGGCAC 5390
CGTGTCTCTTCCCGCGCTCAATTCCAGTTATCTACGACGATGAGCATGTCTATGATGCAGCGTGTAGCA 5460
GTACATCAGCTAGTCAATCTTCGAAACGATCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGA 5530
CA TCGCTGGAGAAATCAGTTCTGATCGTTAACGGGAGCTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGC 5600
5610 5620 5630 5640 5650 5660 5670
AAGGTCAGTGGAAAAAG TGGACTGGAAGATGCTGGATGAAGCTGTTTCCAGTG TCAAGGACATATAT 5670
CTAAATGGACCCAGCTCTACCTCGGAC TAAGCACTGAGTCCA TCCATGGCTACAGCATCAGCCAGCT 5740
GAACGAGTGTGGATGCAGAGCCCCCGAGATGCCCTCT TGGCGTGGAGGTG TCAATAACATATCAGTC 5810
TCCCTCAAAGGTCTGAAGGAGAAA TGGCTGACAGCGCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGA 5880
TGCAGCACTACA TAAGCTCTCTGCTGAAGCACCGGCGCTCGTCTCTCGGCCCCAGCGGCACGGGCA 5950
5960 5970 5980 5990 6000 6010 6020
GACCTACCTGACCAATCGCTTGGCCGAGTACC TGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGGATC 6020
GTCAGCACC TCAACATGSCACAGCAGTCTTCAAGGATCTGCAACTGTATCTT TCCAACC TAGCCAACC 6090
AGATAGACCGGGAAACAGGAATTGGGGAT TGGCCCTCGTGAATCTATTGGATGACC TGAG TGAAGCAGG 6160
CTCCATCAGTGAGTTGGTCAATGGGCCCC TACCTGCAAGTATCATAAATGTCCCTATAT TATAGG TACC 6230
ACCAATCAGCTGTAAAAATGACACCCCAACCATGGCT TGCAC TGGAGCTTCAGGATGTTGACCTTCTCCA 6300
6310 6320 6330 6340 6350 6360 6370
ACAACGTGGAGCCAGCCAATGGCTCTCTG TCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCA 6370
CATCAATGCCAACAAAGGAGAGTGTCTCGGGTGTCTGACTGGGTACCCAAGCTGTGGTATCATCTCCAC 6440
AGCTTCTCTGAGTACAG 6510
GCATTGAGGACTTCCGGACCTGGTTCA TGGAGCTGTGGAACAACCTATCAT TCCCTATCTACAGGAAGG 6580
AGCCAAGGATGGGATTAAGG TCCA TGGACAGAAAGCTGCTTGGGAGGAGCCAG TGAATGGGTCCGGGAC 6650
6660 6670 6680 6690 6700 6710 6720
ACACTTCCCTGGCCATCAGCCCAACAGACCAATCAAAAGCTGTACCACCTGCCCCACCCACCGTGGGCC 6720
CTCAGCATTGCCCTCAGCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGAGTCAGA 6790
TCTCTCATGCCCCATGCTGCTGAAACTTCAAGAAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACC 6860
ATCTTGGACCCCAACCTTCAGGCAACACT TGAAGGTTCCGGCAATCACTG TCAACCCCGGACAGCAGAAC 6930
GC TGGCATCAGCTATCT TAGCTCTCTCTCTCCCTCTCTCTTTCAGAGCAC TGGCTCTCCAGCCCGAGG 7000
7010 7020 7030 7040 7050 7060 7070
AGCAGAACAGGAGGAGGAGGAGATGAAAGAGGAGGAGGACAGG TCTTGGTGTGTACCTTTGAGAAC TCT 7070
CTAGGAAACGAA TGGTGGGGTGGGCT TGGGAACCTGTGCCCC TAAACACATTTACTGGCTTCTCTAAT 7140
GACT TGGGGAAAGATGATTCTGGGTCTTTCCTTGAC TCTTGGTTCAATTACAAC TCTGGGCTTT 7210
CTGGGGAGGGGTTCAGAAACATCAAAACACTGCAAGCAGTTCCCCGGAATTCAGCTTGGAC TTAACAGG 7280
CTGAACCTGCTCAAAAGAAAGCCGAATTCAGCACTGACCTGCCCTCCCCATGGTAT TGA TATCTGAGCTCCGC 7350

Page 4

Monday, 1 December 1997 14:12
fig pCB501

Page 5

```

      9810      9820      9830      9840      9850      9860      9870
      +-----+-----+-----+-----+-----+-----+
      AGCACC GCCTACATACC TCGCTCTGCTAATCC TGT TACCAGTGGCTGCTGCCAGTGCCGATAAGTCGTGT 9870
      CT TACC CGGTGGACTCAAGACGATAG TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGG TCGT 9840
      GCACACAGCCCGAGC TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT TACGATTGAGAAAG 10010
      CCGCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAAGCGGCAGGGTCGTAACAGGAGAGCGC 10030
      ACGAGGGAGCTTCCAGGGGGAAACGCC TGGTATCTTTATAGTCCTGTGGGTTTCGCCACCCTGACTTG 10150
      10160      10170      10180      10190      10200      10210      10220
      +-----+-----+-----+-----+-----+-----+
      ACGTCGAGTTTTTG TGA TGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGGCGGCC TTTT 10220
      ACGGTTCC TGGCC TTTTGCTGCCCTTTTGCTCACA TGTCTCTTCTTCC TCGGTTATCCCCTGATTCTGTGGAT 10290
      AACCGTATTACCGCC TTTGAGT GAGCTGATACCGCTCGCCGCAGCCGAACGACCGA TCGCAGCGAGTCAG 10360
      TSAGCGAGGAAGCGGAAGAGCGCCCAATACGCCAAACCGCCTCTCCCGCGCGTTGGCCGATTCAATTAATG 10430
      CAGC TGGCAGGACAGGTTTCCCGACTGAAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTC 10500
      10510      10520      10530      10540      10550      10560      10570
      +-----+-----+-----+-----+-----+-----+
      ACTCATTAGGCACCCCAAGGC TTACAC TTTATGCTTCCGGCTCGTATGTTGTGTGGAA TTTG TGAAGCGGA T 10570
      AACAA TTTACACAGGAAACAGC T 10594

```

Friday, 28 November 1997 11:36
pTB115

Fig 58

Page

Created: Friday, 28 November 1997 11:02

10 20 30 40 50 60 70
ATGACCATGATTACGCCAAGCTTGCATGCCGTCAGGTCGACTctagAAATGFAAACCTGTCATTTCTGTG 70
TATTTACGACCCGAGAGAGCACCATAAACAGCTACAACAAGAGCACGAACGTCGGCTCCAACAGAAAA 140
GCAGCAAAACGGTGGTACTCACCACATGCTCTGATAATTGCCATTTCTTTCGATTTCTGACTTTCAATGT 210
TGATATGATAACCTAAAAATCTGCCATTCAAAATGAAAACCTTAAAGTTTAAAGAGGCTTCAACGACCC 280
CGCGACTTCCACCCATCATCGTTTTCCCTTCTATCTTATTCTAAATTTTTGTATTTGTGTTCAGTTTGC 350
360 370 380 390 400 410 420
TATATCCATCATCACCCTTTTACCGTTTAAATCTCTCTATAGTATTGTTCCGGTGGTTTCAAGATGAAACG 420
TTTGTGTGTAGCATTTTGAACACAGCGGAGAAAAATGCAGAAACATCATCTCCCAATGCAATTAAGTGC 490
ACTCATTGTGATGACTCCAAACCCCTCGCCCCACCATTTGTCTCTTCACAAATCAAGGACCAATGCAA 560
AAATGGCTTTTTGTGTGGATCTTGATCTGCTCTCTTCTCTCTTCTTTCGTTCTCTCGGACCACTCAT 630
TATTCGAATGCCATGCAATTTGTGATGATGCGCAGCATGGTCGCACACACTACAACAGATATGATGGTGG 700
710 720 730 740 750 760 770
CTGTGCGTGGTGCAGAGCTCATGAATAAGTGATGGGGCACCGAGATGTCGACTCCCTCCCATTAATCTCTT 770
GAGTCGCATCTTTGTCTTGGTCCGTTGTCTGATCGAGCCTCAAGGCAGCCGTTCCGCCAGTGATCGCCCTG 840
CTAAGAAGTTGTAGGTGTAGAGGAAGAACGTCGGGTCCAAATTTCAAGGAGGAGGAGGAGGAGGAGGAG 910
tcccgGGGATTGGCCAAAGGACCCAAACGATGTTTTCGAATGATACTAACATAACAAGAAACATTTTCAG 980
GAGGACCCTTGGGctaggaactagtgagatcagagctctcccatATGACGACGTCAAAATGTAGAAATGATACC 1050
1060 1070 1080 1090 1100 1110 1120
AATCTACACGGATTGGGCCAATCGGCACCTTTTCGAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTC 1120
AATGATTTTCGCGACTATCGACTGGTTCTCAGCTTATTAATGTCGATCGTTCCGATCAACGAATCTCCG 1190
CTGCATTACGAAACGTTGGGCAAAAAACACATCGAACCTGGATGGCTCGAAACGTTCTCGACTACCT 1260
GAAAAATCTGGGCTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACAGCGGAAACATGGGTGCAAT 1330
CTCCAGCGCTCTCTCTGCTCTCCACCTACAAGCAGAAGCTTCGGCAACTGAAAAAAGATCAGAAGAAAT 1400
1410 1420 1430 1440 1450 1460 1470
TGGAGCAACTACCCACATCCATATGCCACCCGCGGTTTCTAAATTACCTCGCCACGTTGTGCCACGTC 1470
AGCAACCGCTTCAGCAACATACCCAAATTCCAACTTTCACCAAAATGTCAACATCCAGGCTTCAGACTCCA 1540
CAGTCAAGAAATATCGAAATATGATTCATCAAGATTGGTATCAAGCCAAAGACGCTGGACTTAAACCAC 1610
CCTCATCATCAACCACTTCAATCAATAATACAAATTCATTCCGTCCTCGAGCCGCTCGAGTGGCAATTA 1680
TAATGTTGGCTCGACGATATCCACATCTGCCAAGAGCTTAGAATCATCATCAACGACAGCCTATTTCG 1750
1760 1770 1780 1790 1800 1810 1820
AATCTAAACCGACCTACCTCCCAACCTCAAAACCTTCTAGACCACAAACCCAGCTAGTTCCTGTGTGTCTA 1820
CAACATCAAAATCGGAAGCTCAAGCTAGCCGCTCCGAAAGCGGTGAGCACCCCAAAACATGCTTCTGT 1890
GAAGACTATTGGAGCAAAACACAGCCCTAATACAGCGGTGGTGGTGGTGGGATGCTGAAATTAAG 1960
TATTCAGTAGCAAAACCCAATCTCTCTCATCGAATAGCCCAACCTACAGAAAGCGCGGCGGCGGTGC 2030
CTCAACAAACATTTGTTCGAAATCTCTGCCCCAGTGAAGAGTGGCTTGAAGCGCGCGACCAAGTAACT 2100
2110 2120 2130 2140 2150 2160 2170
GGGAAGTGGCACGTCATGTGGAAGCTTTGTACGCCAAAGTTTCTACCGTAAACCGGACGCCCCAATC 2170
ATATCTCAACAAGCTCGAAACGATGCTCAAGAGCAGTGAAGAAGAGTCGGATACGCTGGATCAACA 2240
GCAGCTCGCAACGTCATCAATGAGGAGGTTCCTAAGCAATGCAATCCACATCTTCCAAAGAGTCAAC 2310
GTCAGACGAAAGCTCTCGCTCATCAGGATCTTACCTTAAAGCCTCCATCGTGACAGCTATCAGACAG 2380
CCGATAGCCGCAACACCGGTTCTCCAAATATATCAACAAGCTTGTGAGGAAACCAACCACTGGCAG 2450

Friday, 28 November 1997 11:36
pTB115

Page

```

      2460      2470      2480      2490      2500      2510      2520
      .....
TGAAAGCAGTGAAAAGCACAGCGAAAAAAGATCCACCCTCCAGCTGTTCCGCCACGTCACACCCAGCCAAAC 2520
AATCCGAGTTGTTAGTCCAACTATGGCACATAAGAAGTTGACAAATGACCCCGTGATATCTGAAAAACCA 2530
GAACCTGAAAAGCTCCAATCAATGAGCATCGACACGACGGACGTTCCACCGCTCCACCTGCAAAAATCAG 2540
TTGTCACCTGAAAAAGACCTCAATCCGACACCCACCAACGTACGATGTTCTTCTAAAACAAGGAAAAAT 2550
CAGATCGCCCTGCAAGTCGTTTGGATATGAGCAGTCCTCCCGCTCTGAAGACTCCATTGTGGCTCATGCG 2560
      2810      2820      2830      2840      2850      2860      2870
      .....
TCGGCTCAGGTGACTCCGCCGACAAAAACTTCTGGTAATCATTCCGCTGGAGAGAAGGAAGGAAAGAATA 2870
AGACATCAGAATCCAGCGGCTACACCTCTGACGCGGGTGTTGCGATGTCGCGCCAAAATGAGGGAGAAAGCT 2880
GAAAGAAATACGATGACATGACCTGTCGAGGACAGAACGGCTATCCTGACAACTTCGAAGACAGTTCTCTCC 2890
TTGTCGTCCTGGAAATCCGATAACAAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACA 2900
TGGCAACAGTCGGCTCCAAACATAGCGACCTATCCACCTTTGTTGCCATCCACGCTCTCTCTCCCTCAAA 2910
      3180      3190      3180      3190      3200      3210      3220
      .....
GCCCGAGTCCCCAGTCGGTCTCCACATCAGTCGATTCTCGATCTCGAGCAGAACAGGAGAATGTGTAC 3220
AAACCTCTGTCGCGCAGTCCGGAACGAGCCAAACGTCGGCGCGCTGCCACCTCAACCTTCGGGACAACTCTCC 3230
TAAGATCCCCGGGATCTCACTCTGATCTCTGACACCTTATCAGTGTGAGCTGATTAAGGACACAAAGTCTAT 3240
CCACTCAGAGACTAGTCGACGACCTTCTTCAAAAAACCAAGCTATTTCAGGCCAAATTCATTCACTTGAT 3250
CGTAAATGCCACCTTCAAGAGTTCACTCCACCGAGCACAGAATGGCGGCTCTCTTGAAGCCGAGACGGG 3260
      3510      3520      3530      3540      3550      3560      3570
      .....
TGCCGAACCTCGATGTCGAAATATGATCTCTCAGGATCCTACTCGGCGCGTTCCCGAGGAGGAAGCTCTAC 3570
TGGTATCTATGGAGAGACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCGACATTCTGCCAAA 3580
AGTGAGATGGGATCCCAACTATCACTGGCTAGCAGCAGACATATGGATCTCTCAATGAGAAGTACGAAC 3590
ATGCTATTTCGGGACATGGCAGCTGACTTGGAGTGTACAAAGAACACTGTCGACTCACTAACCAAGAAACA 3600
GGAGAACATATGGAGCACTGCTTGAATCTCTTGAAGCAAAAGCTTAGAAAACCACTCAACCAATGATTCGA 3610
      3860      3870      3880      3890      3900      3910      3920
      .....
TCCAACCTTGAAAGCTGAAAGAGGCAATACGATTACGGCAGGACATTGCTCATTGAGGGATATTAGCAATC 3920
ATCTTGCATCCAACTCAGCTCATGCTAACGAAGCGCGTGGTGAGCTTCTTCGTCAACCATCTCTGGAACT 3930
AGTTGCATCCCACTGATCTGATGTCATCTGTCGTCGAAAAAGCAGCAAGCAGGAGAGATCAGCTTGAGC 3940
TCGTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGTTCACCAAGAAGAAGAACAGA 3950
ACTACGACGAAGCACATATGCCATCAATTTCGGGATCTCAAGGAACCTCTGACAAACATGATGTGATTGA 3960
      4210      4220      4230      4240      4250      4260      4270
      .....
GTTGAAGCAAGAGCTCAAGAAGCGGAAGTGCACCTTACGAAGTCCGCTTGACATCTGGATCTGTGCC 4270
CGCGAAGTTGATGTTCTGAGGGAGACAGTGAAACAAGTTGAAAACCGAGAACAAGCAATTAAGAAAGAAAG 4280
TGGACAAACTCACCAACGGTCCAGCCACCTGTCGCTCTCTCCCGCGCTCAATTCCAGTTATCTACGACGA 4290
TGAGCACTGCTATGATCAGCTGTAGGAGTACATCAGTCAATCTTCGAAACGATCTCTGGCTGCT 4300
AACCTCAATCAAGGTTACTGTAAACGTGGACATCCGTGGAGAAATCAGTTCGATCTCTTAACCGGACAAAG 4310
      4580      4590      4600      4610      4620
      .....
AGATAATCGTAGGATATCTTGCATGTCACACAGTCAGTCATGCTGGAAAGACATTTGATGTTCTTATCTCT 4620
AGGACTATTTGAAGCTATCTTATCTTCAAGAAATGATGTCGAGGATCAACTTGGAAATCATGCTCTGATCTCT 4630
ATCCTTGGCTATCAAACTTGTGAACTTTCAGCGCTCATTTGGAGACTCCACAACCAATGATACCAAGCACT 4640
CAACTGACATTTCTTCTTCAACTTCAATCCGAATGTTTCATGACCGGTGCCGACAGAGCTGGCTGAGA 4650
CAGTCTGGTCTCTGATATGCTTCTTCAAGCAAAATGATTCTCCAACCTCGTCAAGTCAATTTGACAGAG 4660

```

Friday, 20 November 1997 11:37
pTB115

Page

4910 4920 4930 4940 4950 4960 4970
AGACGTC TGGTGT TAGCTGG AGCAACTGGAATG GAAAGAGCAAAC TGGCGAAGAC CC TGGC TGGTTATG 4970
TATCTAT TCGAACAAA TCAATCCGAAGATAGTATTGTTAATATCAGCATTCTTGAA AACAATAAAGAAGA 5010
ATTGCTTCAAGTGAACGAC GCCTGGAAAAGATCTTGAGAAGCAAAGAATCA TCGA TCGTAATTCTAGAT 5110
AATATCCCAAAGAAATCGAATTGCAATTTGTTGATCCG TTT TCGAAA TGTCCAC TCAAAAACACGAAG 5130
GTCCATTTGTAG TATGACAG TCAACCGATA TCAAAATCCCTGAGCTTCAAATTCAACACAA TTTCAAAA 5250
5260 5270 5280 5290 5300 5310 5320
GTCAGTAATGTGGAATCGTCTCGAAGGATTCACTACGT TACCTCCGACGACGGGCGGTAGAGCATGAG 5320
TATCTCTAACTGTACAGATGCCATCAGAGCTCTTCAAATCATTGACTTCTTCCC AATAGCTCTTCAAG 5390
CGGTCAATAATTTTAT TGAAGAAAACGAATCTGTTGATGTGACAGTTGGTCCAAGAGCA TGT TGAAC TG 5480
TCCTCTAAC TGTGATGGATCCCGTGAATGCTTCATTGATTGTGGAATGAGAATCTCATTCCATA TTTG 5530
GAACGTGTTGCTAGAGATGGCAAAAACCTTCCGGTCCGTGCAC TCTCTCGAGGA TCCACCGACATCG 5600
5610 5620 5630 5640 5650 5660 5670
TCTCTAAAAATGGCCGTGGTTGATGGTGA AAAACCCGGAGAATGTGCTCAAACGTCTTCAACTCCAGA 5670
CC TCGTCCCGTCACTTGGCAACTCAT TCCGGACAACAC TCAATCCCTCGAGTCTGTGATCCAAATTGCAT 5740
GCTACCAAGCA TCAAGACATCGACAACAT T TGAACAGAAGAC TCTAA TCTCTCGCTCTCCCCCGCT 5810
TTCTTATCTTCTACCGGTACCATGGTATTGATATCTGAGCTCCGCATCGGCCCGTGTGATCAGATCGC 5880
CA TCTCGCGCCCGTGCCTCTGACTTCTAAGTCCAATTACTCTTCAACATCCCTACATGCTCTTCTCCCT 5950
5960 5970 5980 5990 6000 6010 6020
GTGCTCCCAACCCCTATT TTTGTTATTATCAAAAAA ACTTCTTCTTAATTTCTTTG TTTTTTAGCTTCTT 6020
TTAAGTCACTCTAACAATGAAATTTGTGATGATCAAAAATAGAATTAATTCGTAA TAAAAAGTCGAAAA 6090
AAAT TGTGCTCCCTCCCCCAT TAA TAA TAA TCTATCCCAAAATCTACACAATGT TCTGTGTACACTTC 6160
TTATGTTTTTTTACTTCTGATAAATTTTTTTGAAACATCATAGAAAAACCGCACAAAA TACCTTA 6230
TCATA TGTACGTTTCACTTATGACCGCAATTTTTATTTCTTGGCAGTCTTGGGCTCTCTCA TGAAGTCA 6300
6310 6320 6330 6340 6350 6360 6370
AATCATGCTCATCGTGAAAAAGTT TGGAGTATTTTGGAAATTTTCAATCAAGTCAAGT TATGAAAT 6370
TAATTTTCTGCTTTTCTTTTGGGGGTTTCCCTATTGTTTGTCAAGAGTTTCGAGGACGGCGTTTTT 6440
CTTGCTAAAAATCACAAGTAT TGA TGAACAGGATGCAAGAAAGATCGGAACAAGGTT TGGGTTTGAGGC TC 6510
AGTGAAGGTGAG TGAAGTTGATAATTTGAAAG TGGAGTGTCTATGGGGTTT TGGCTTAAATGAC 6580
AGAAATACATTCCTCAATATACCAACATAACTGTTTCTAC TAT TCGGCCGTACGGGCCCTTTCTGCTCTGC 6650
6660 6670 6680 6690 6700 6710 6720
GGGTTTCGG TGA TGAACGGTGAAAACTCTGACACATGCAGCTCCCGGAGACGGTCA CAGCTTGTCTGTAA 6720
GGGATGCCGGGAGCAGACAAGCCCTCAGGGCGCG TCAACCGGTGTTGGCGGGTCTCGGGCTGGCTTA 6790
ACTATCCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATA CCGCACAGATGGCT 6860
AAGSAGAAATACCGCA TCAAGGCGCCCTTAAGGGCCTCGTGATACGCC TATTTTA TAGGTTAATGTCA 6930
GATAATAATGTTTTCTAGACCTCAGTGGCACTTTTCGGGGAATGTGCGCGGAACCCCTATTTGTTTA 7000
7010 7020 7030 7040 7050 7060 7070
TTTTTCTAAATACATTCAAAATATG TATCCGCTCATGAGACAATAACCTTGATAAA TGTCTCAATAA TATT 7070
GAAAAAGGAAGATATGAGTAT TCAACATTTCCGTG TCGCCCTTATTCCTTT TTTGCGGCA TTTGCTT 7140
TCCTGTTTTGCTCACCAGAAACCTGG TGAAGTAAAGATGCTGAAGATCAG TGGGTGCACGAGTG 7210
GGTACATCGAAGTGGATCTCAACAGCGGTAAGATCTTGAAGT TTTCCGCCCGAAGAAGCTTTTCCAA 7280
TGATGAGCAGTTTAAAGTTCTGCTAT TGGCGCGGTATTATCCCG TAT TGAACGCC GGGCAAGAGCAATCT 7350

Friday, 28 November 1997 11:37
pTE115

Page

```

      7360      7370      7380      7390      7400      7410      7420
      .....
CGGTCGCGCGCAACACATATTCTCAGAAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACG /420
GATGGCATGACAGTAAGAGAAATAAGCAGTGTGCGCATAACCATGAGTGATAACAC TGCGGCCAAC TAC 7430
TTCTGACAACCATCGGAGGACCGAAGGAGCTAACCCTTTTTTGCACAACATGGGGGATCATGTAACTCG /580
CCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGAAGCCGTGTA 7630
GCAATGGCAACAACGTTGCGCAAAC TATTAAC TGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAA 7700
      7710      7720      7730      7740      7750      7760      7770
      .....
TAGACTGGATGGAGGCGGATAAAGTTGCAGGACCCTTCTCGCCTCGGCCCTTCCGGCTGGCTGGTTTTAT //10
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAG /840
CCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCG 7910
CTGAGATAGGTGCCTCACTGATTAAAGCATTGGTAACGCTCAGACCAAGTTTACTCAATAATACCTTAGAT 7980
TGATTTAAACTTCATTTTTTAATTTAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAA 8050
      8060      8070      8080      8090      8100      8110      8120
      .....
ATCCCTTAACGTCAGTTTTCGTTCCAC TGAGCGTCAGACCCCGTAGAAAAGATCAAGGATCTCTCTAGAG 8120
ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACACCGCTACCAAGCGGTGGTTTTGTTT 8150
GCGGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAAATACT 8280
GTCTTCTAGTGTAGCGGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGGCTACATACCTCGCTC 8330
TGCTAATCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACCAAGACG 8400
      8410      8420      8430      8440      8450      8460      8470
      .....
AAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGCGA 8470
ACGACCTACACCGAACTGAGATACC TACAGCGTGAGCATTGAGAAAGCGCCACGCTTCCCGAAGGGAGAA 8540
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCAGGAGGAGCTTCCAGGGGAA 8610
CGCCTGGTATCTTTA TAGTCCGTGTCGGGTTTCCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCG 8680
TCAGGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCTTTTACGGTTCTCGCCTTTTCTGCGC 8750
      8760      8770      8780      8790      8800      8810      8820
      .....
CTTTTGCTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGT 8820
AGCTGATACCGCTCGCCGCGAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGCAAGCGGAAGAGCGC 8890
CCAATACGCAAAACCGCTCTCCCGCGCGTTGGCCGATTCATTAATGCACTGGCAACACAGGTTTCCCG 8960
AC TGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCAACCCAGGCTTT 9030
ACACCTTATGCTTCCGGCTCGTATGTTGTGTGGAATTTGAGCGGATAACAATTTACACAGGAACACG 9100
      9110      9120      9130      9140      9150      9160      9170
      .....
T 9101

```


Friday, 28 November 1997 11:00
pPD95-75

Page

Created: Friday, 28 November 1997 10:58

```

10      20      30      40      50      60      70
AAGCTTGCAATGCCCTGCAGGTGCGACTCTAGAGGAATCCCGGGATGGCCAAAGGACC CAAAGgtatgttctc 70
gaatgatactaacatgaatagaaatttccagGAGGACCCCTTGGAGGGTACCGGTAGAAAAAATGAGTA 140
AAGGAGAAGAACTTTTCACGGAGTTGTCCCAATTCCTGTTGAATTAGATGGTGAATGTTAAATGGGCACAA 210
ATTTTCTGTCACTGGAGAGGGTGAAGG TGA TGC AACA TACGGAAAACTTACCCTTAATTTAATTTGCACT 280
AC TGGAAAACTACCTGTTCCATGGgttaagttaaactatataactaaactaaacttgattatttaagt 350
360      370      380      390      400      410      420
ttcagCCAACAC TGTCACTACTTTCTgTTATGGTGTTC AATGCTTcTCgAGA TACCAGATCATATCAA 420
ACgGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTCAAAGAT 480
GACGGGAAC TACAAGACACgttaagttaaaccagttcggtaactaaactaaactaaactaaactaaactaaact 560
GTGCTGAAGTCAAGTTTGAAGGTGA TACCC TGT TGAATAGAGTTCGAGTTAAAAGGTATTGATTTTAAAGA 630
AGATGGAAACA TCTTGGACACAAATTGGAATACAAC TATAACTCACACAATG TATACATCA TCGCAGAC 700
710      720      730      740      750      760      770
AAACAAAAGAAATGGAATCAAAAGTTgtaagttaaacttggaacttaactaaactaaaggatatactttaagt 770
ttcagAAC TCAAAAATGAGACACAACA TGAAGATGGAAAGCGTCAACTAGCAGACCATTAATCAACAAAA 840
TACTCCAAATGGGCGATGGCCCTGTCTCTTTTACCAGACAACCATTACCTGTCCACACAAATCGCCCTTCG 910
AAAGATCCCAACGAAAAGAGAGACCATGGTCTCTCTTGAGTTTGTAAACAGCTGCTGGGATTACACATG 980
GCATGGATGAAC TATACAAATAGCAATTCGTAGAATTCCAAC TGAGCGCGGCTGCTACCATTAACCACTT 1050
1060      1070      1080      1090      1100      1110      1120
GTCTGGTGTCAAAAATAATAGGGGGCGCTGTATCAGAGttaggttaaacttgagttctactaaactaaaga 1120
gtaatatttaaattttcagCA TCTGCGCCCGTGCCTCTGACTCTAAGTCCAATTACTCTTCAACATCC 1190
CTACATGCTCTTTCTCCCTGTGCTCCACCCCTATTTTGTATTATCAAAAAAATCTCTCTTAATTT 1260
CTTTTGTCTTTTGAAGTCACTCTAACAATGAAATGTGTAGATTCAAAAATAGAAATGAAT 1330
CGTAA TAAAAAGTCAAAAAAATTTGCTCCCTCCCCCA TTAATAATAATCTATCCCCAAATCTACAC 1400
1410      1420      1430      1440      1450      1460      1470
AATGTTCTGTGTACAC TCTTATGTTTTTTTACTTCTGATAAAATTTTTTTGAAACA TATAGAAAAAA 1470
CCGCACACAAAAATACCTTATCATATGTTACGTTTCAGTTTATGACCGCAATTTTATTCTCTCGCAGCTC 1540
TGGGCTCTCATGACGTC AATCATGCTCATCGTGAAAAAGTTTGGAGTATTTTGGAAATTTTCAATC 1610
AAGTGAAGTTTATGAAATTAATTTCTCTGCTTTTGTCTTTTGGGGTTTCCCTATTGTTTGTCAAGAG 1680
TTTTCGAGGACGGCTTTTCTTCTTAAATCACAAGTATTGTATGAGCACGATGCAAGAAAGATCGGAAGA 1750
1760      1770      1780      1790      1800      1810      1820
AGGTTTGGGTTTGGAGGCTCAGTGGAGGTGAGTAGAAGTTGATAATTTGAAAG TGGAGTAGTGTCTATGG 1820
GGTTTTTGGCTTAAATGACAGAAATACATTCCCAATATACCAAAACATAACTGTTTCTTACTAGTCGGCCCT 1890
ACGGGCCCTTCTGTC TCGCGCGTTTCCGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGAC 1960
GGTCACAGCTGTCTGTAAAGCGATGCCCGGAGCAGACAAGCCCGTCAGGGCGCTCAGCGGGTGTGGC 2030
GGGTGTGGGGG TGGCTTAACTATGCGGCA TCAAGAGCAGATGTACTGAGAGTGCACCATATCGGGTGTG 2100
2110      2120      2130      2140      2150      2160      2170
AAATACCGCACAGATGCGTAAAGSASAAAATACCGCATCAGGCGGCTTAAGGGCTTCG TGA TACGCCAT 2170
TTTTATAGGTTAA TGTATGATAA TAAATGTTTCTTACAGCTCAGTGGCAC TTTTCGCGCAAAATG TCG 2240
CGGAACCCCTAT TGTATTATTTCTTAAATACATTCAAATATG TATCCGCTCA TGAAGACAA TAACTCTGA 2310
TAAATGCTTCAATAA TATTGAAAAAGGAAGAGTATGAGTATTCAACA TTTTCGTGTCGCCCTTATTCCT 2380
TTTTTGGGGCATTTCTCTCTCTGTCTTTTGTCTACCCAGAAACGCTGGTGAAGT TAAAGA TGTGAAGA 2450

```

Friday, 28 November 1997 11:00
pPD95-75

Page

2460 2470 2480 2490 2500 2510 2520
TCAGTTGGGTCACGAGTGGGTTACATCGAAC TGGATCTCAACAGCGGTAAGA TCC TCGAGAGTTTTTCGC 2520
CCCGAAGAAGCTTTTCC AATGATGAGCACTTTTAAAGTCTTGCTATGTGGCGCGS A TATCCCGTATTG 2530
ACGCCGGGCAAGAGCAAC TCGGTCGCCGCATACAC TATCTCAGAATGAC TGGT T GAGTACTCACCAGT 2660
CACAGAAAAGCATCTTACGGATGGCA T GACAGTAAGAGAA TATGCAAGTGCTGCCA TAACCATGAG TGA T 2730
AACACTGCGGCCAAGCTTAC T T C T GACAACGATCGGAGGACCGAAGGAGCTAACCG T T T T T GGCACAACA 2800
2810 2820 2830 2840 2850 2860 2870
TGGGGGATCATGTAAGTCCGCT T GATCGTTGGGAACCGGAGCTGAATGAAGCCA TACCAAACGACGAGCG 2870
TGACACCACGATGCC TGTAGCAATGGCAAC AACGTTGCGCAAACTATTAACTGGCGAACTAC T TAC T C T A 2940
GCTTCCCGGCAACAA T T AATAGACTGGATGGAGGCGGATAAAGTTGCGAGGACCAC T TCTGCGCTCGGCC 3010
T TCCGGC TGGCTGGTTTATTGC TGA TAAATCTGGAGCGGTTGAGCG TGGTCTCGCGGTATCAT TGCAGC 3080
ACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAGGACGGGGAG T GAGGCAAC TATGGAT 3150
3160 3170 3180 3190 3200 3210 3220
GAACGAAATAGACAGATCGCTGAGATAGG T GCTCACTGATTAAGCAT TGG T AACT TGTACAGCAAGTTT 3220
ACTCATATATACTTTAGATTGA T T TAAACTTCATTTTTAATTTAAAGGATCTAGGTGAAGA TCC T T T T 3230
TGATAATCTCATGACCAAAA TCCCTTAACGTGAGTTTTCTG T TCCACTGAGCGTCA TACCCCGTAGAAAAG 3360
ATCAAAGGATCTTCT T T GAGATCTTTTTTTCTGCGCGTAATCTGCTGC T T GCAAAACAAAAAACCACCCG 3430
TACCAGCGGTGG T T T G T T T GCGGGATCAAGAGC TACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG 3500
3510 3520 3530 3540 3550 3560 3570
AGCGCAGATACCAATAC T G TCTTCTAGTGTAGCCGTAGTTAGGCCACCAC T TCAAGAAC TCTGTAGCA 3570
CGGCTACATACCTCGC T C T GCTAATCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTA 3640
CGGGTTGGACTCAAGACGA T AGTTACCGGATAAGGCGCAGC3GTCCGGCTGAACGCGGGG T TCGTGCAC 3710
ACAGCCAGCTTGGAGCGAAGGACCTACACCGAAGTGAAGATACCTACAGCGTGAGCATTGAGAAAAGCC 3780
ACGC T TCCCAAGGGAGAAAAGCGGACAGGTATCCGG T AAGCGGCAGGCTCGGAACAGGAGAGCGCACGA 3850
3860 3870 3880 3890 3900 3910 3920
GGGAGC T TCCAGGGGGAAACGCCCTGGTATC T TATAGTCTG TCGGGTTTCGCCAAGCTC T GAC T T GAGCG 3920
TGA T T T T TGTGATGCTG T T CAGGGGGGCGSAGCC TATGAAAAACCCAGCAACGCGGCC T T T T TACGG 3990
TTCTTGGCC T T T T GCTGGCC T T T T GCTCACATGTTCT T TCTGCGTTA TCCCTG T TCTGTGGATAACC 4060
GTATTACCCCT T T T GAGTGACC TGA TACCSCTCCGCGCAGCCGAAACGACCAGCGCAGCGAGTCAGTGAG 4130
CGAGGAAGCGGAAGAGCGGCCAATACGCAACCGCCTCTCCCGCGCGGTTGGCCGA T TCA TTAATGCAGC 4200
4210 4220 4230 4240 4250 4260 4270
TGGCACGACAGGTTTCCCGAC TGGAAAGCGCGGCAAGTGAGCGCAACCGCAATTAATGTGAGT TACCTCAC T C 4270
AT TAGGCACCCAGGCT T TACACTTTATGCT TCCGGCTCGTATG T T G T T GGAAT TGTAGCGGATACCA 4340
ATTTACACAGGAAACAGCTATGACCA T GATTACGCCAAGCT T gtaagt ttaaacatgata: ltaactaacta 4410
gctatctcat: taaat t ttaagAGCTTAAAAA TGGCTGAAATCACTCACACGA TGGATACGC TAAACA 4480
CTTGGAAA TGAAT T 4490

Friday, 28 November 1997 13:10
pBS KS/X18

Page

Created: Friday, 28 November 1997 12:02

10 20 30 40 50 60 70
CTAAATTGTAAGCGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCACTCATTTTTTAAC 70
CAATAGGCCGAAAICGGCAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGTTGAGTGTGTTTC 140
CAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGICAAAGGGCGAAAAACCGTCATCA 210
GGCGGATGSCCCACTACGIGAACCAICACCCIAATCAAGTTTTTGGGGTGGAGGTSCCGTAAAGCACTA 280
AATCEGAACCTAAAGGGAGCCCGGATTTAGAGCTTGACGGGAAAGCCGGCGAAAGTGGCGAGAAAGG 350
360 370 380 390 400 410 420
AAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGGTGGCGGTAAACCAC 420
CACACCGCGCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCAICGCGCAICAGGGTGGCGAACIGTGG 490
GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCIGGCGAAAGGGGGATGTGCTGCAAGGCCAT 560
TAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAACGACGGCCAGTJAGCGCGCGTAAATA 630
CGACTCATATAGGGCGAATGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACGAGTGGATCCCCCGGG 700
710 720 730 740 750 760 770
CTCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGGATCAGAAAGAAATGGAGCAACTAC 770
CCACATCCATTATGCCACCGCGGGTTCCTAAGTGAGTTTAAATTTGAGTTTACGACTACAAAAATGTGT 840
CTTTAATAACTATCTTCGACTTGAGTCTATTCTGTATCACTAGTTGTGAGTGTATTTCATTGAGAAAA 910
TATTAAGGAACATTTATTTACTTTGCCTATTTCGCCCTAACITTGATTTAGTTTTTSGAICAACTAGATC 980
TTACAAAACCTTGCAATACAATTTCCAITTTTCAGATTACCTCGCCACGTGTGCCACGTGAGCAACCGCTT 1050
1060 1070 1080 1090 1100 1110 1120
CAGCAACTAACCCAAATTCCTAACTTTCCACAAATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGAA 1120
ATCAGAAATTTGTAAGAATTTTATTTTGTAGCTCAAACCTGTATAAAATGCCCAAGAAAGATGATATA 1190
AATGTAGTTTTTTGCAAAACTTCCACCTTTATTGCTTAATATGACGGCTTATATCTCAAATTTCTTGA 1260
GTTTTATCAAAAAATTTCCACTATACAAATGTAGAAAAGTATTTGCACAAATTTTGTCAGTTGACAGC 1330
TTTGTAATAGATCCAAATGGAACCTAGATACAAGCTGTCTAAAGTGAAGGAGCGCAAGTCTATACTGGA 1400
1410 1420 1430 1440 1450 1460 1470
ATAATGATCTGAAACAAATTTGTGCTATTCCTAAATGTTTAAAGACATGTTTTGAAGATTTTCTCAATTC 1470
GCACTAGTTTCAGAACCTTCCTTTTGTATGAAAAAGTAAAAAAAACCTATTTCAAACCTTCACCGCCAC 1540
CATGTTTCAACTCTTAATTTTATAAAATTTGCAATTTACAAATCGCCCTCCCTTGCCCGAAAGTCC 1610
CACCAAAAICAAATTTCTCGGCCTCATAATGACTTTTAAATGATGTGAGAAAACACAGAGAGGGTAACT 1680
AAATGACAGGGACAGGTTGTCTCTCTCTCCCTCTCTCTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1750
1760 1770 1780 1790 1800 1810 1820
AACAACAATTTTCCAAITTCGTGTGCTATTTGCTTATAAACAIITGTGTGTGGAAGGAACIACACGGG 1820
GAGACGGTCAATTAATTTGAATGAGAGCATGGCAATTACTCTTTCGGAAATTTGATCAAATAAGATAGAGC 1890
CGATGACACTGGCTGGTAGTAGTATGAGTGTAGAAATGCTTTTCATCGTCTCAACTTGCCTCATGCTCT 1960
TCCCCGCTCTCATCACTGACAATTAATGTGGGTTTTATGCGCTCTTTCCTATTCGCGCACIATCTCTG 2030
GGTTACCAACAACTGGAATACATTTTACTACTATTCAAGCCATTATTTTGATATTTAAATTTGTGCAAT 2100
2110 2120 2130 2140 2150 2160 2170
TAGGGATAAACACGACTTTTAAAGTTTATTTAAAAAACGATATTTTCGATTTTAAAAAATCTGAAAG 2170
TTTCAAAAAATCAATAATAATTTCCCTTAAACAAATTTGTATGGCTAAAAATTTATTTCTACTGTGACAAT 2240
CTTTATATGTATCACTGTTTTCATCTCAAAACCTTGAATCCCCCAAGTTATAGGAGGCTCCGTGTCAAA 2310
TTTCCCATGCTATGAATCGGTACGAGGACATATCCAAAAATTAAGCTAGACGGTGTATAATTAATTTGGG 2380
AGCGCTAAATAAGTCAAGGAGTATAGAAATTAATTAAGGACAGATTATCTATCAATTTCAATCTTGA 2450

Friday, 28 November 1997 12:03
pBS KS/X18

Page

2460 2470 2480 2490 2500 2510 2520
ACATTCAGCCAGTTCGTACAAATTTCCATGCTTTTGGCCCATTAATAAACCTCTCACCTCTTCAATCA 2520
TCTCACTCGTATCAAAAAAGTATAGCAAAAGCCGACTCTACTTTTAAAGAGAAGGAGATACTGAGCCA 2590
CAIGGCGGTGACCCCTTTTCATCTCGTCCGTTCCGCTCTCAAAATTCACGCCTATACTAACTCTTCAAAAG 2650
CCATAGACCTCCCTGTTTTCTTCTTCTTCTGACTCGCGCCTATTTTGTGGCTGCTGAAAGCCGCGGA 2720
AAATTTAGTATATTTATGAGCTTATCTTTATGCAATACATAAAAAACGAGGCAATTTAAAAATATTAATA 2800
2810 2820 2830 2840 2850 2860 2870
TAAATGAGGTTGTAGATGTAGATTTGGAAAAGAAGAAAAAACAAACAAATAGGAACCGCCAGATCAAA 2870
ATTCTATTTAAAGGTTTCTAAGATGTTTAGGCAAGATTCGGCTGAACAGAAAACTGAAGTGCCCTCATAA 2940
ATCTAGTGAACGTTTAGATTGAACCTGGAAATCTTAAGCCCTGAACATATAGCCTTATTCTAGATCTTAGT 3010
TGGCATAAGCTCAAGCCCAAGCAGAAATGACTTGCATTTAGTTTAAAGCCCTAGATTGACTTGTCTTCT 3080
AGTCTAATCCAGACTAGATTTCCAAGAGAGTTTCAATTTTAAATGTTTCCAGTTCTTCTTACTTAAAA 3150
3160 3170 3180 3190 3200 3210 3220
TCATAATGAGGTTGTAGTGGTAAATCGTTATCCCTTCTCTCACACTTTCAATTACAGATTTCATCAAAAG 3220
ATTGATATCAAGCCAAAGAGCTCTGGACTTAAACCACCTCATCACTAACCTTCTATCAAAATATTAACAA 3290
ATTCATTCCGTCCTGAGCCGTTCGAGTGGCAATAAATGTTGGCTGAGCAGATATCCACATCTGCGAA 3360
GAGCTTAGGTATCCGATCTCTGCGCTTCTTTTAGAAATATATTAATTCAGAAATCATCATCAACGTA 3430
AGCTCTATTTCGAATCTAAACCGAGCTACCTCCCAACTCCAAAAACCTCTAGACCAAAACCCAGCTAG 3500
3510 3520 3530 3540 3550 3560 3570
TTCGTGTTGCTACAACCTACAAAAATCGGAAGCTCAAAAGCTAGCCGCTCCGAAAGCCCTGAGCACCCCAAA 3570
ACTTGCTTCTGTGAAGACTATTGGAGCAAAACAAAGAGCCGATAACAGCGGTGGTGTGGTGGTGGAAATG 3640
CTGAATATTAAGTTATTCAGTAGCAAAAAACCCATCTTCCTCATCGAATAGCCCAACCTACGAGAAAGC 3710
CGCCGGCGGTGCTCTACAAACAAACCTTTGTCGAAAACTGCTGCCCCAGTGAAAGAGTGGCTGAAGCCGCC 3780
GACCAGTAAGCTGGGAAGTCCCACTCTATGTCGAAGCTTTGTACGGTGAGTATTTGAAATCGGAAATTTG 3850
3860 3870 3880 3890 3900 3910 3920
GAAATGATATTTTAAAAACCTGAAATTTCTACAAAAATAAAATAAAAAAGAGATTTTCTCTCTGATAGTA 3920
TTGCATCCCACTATTTTACCTTGAAGATTTATATCTTGGTTCATATTGAAGATATCAGATAAGAAAAAG 3990
AAATAAAAAATATTTTACAGTTGATAATTTTGTATAGGACCAAAAGACAAAGTGAGATATAAGCTGTCAA 4060
AGTTGATTTTCAAGAAATTTTAAACCCCTAGTTTTCGGAAGCTCTGGGCTCATCTATATTAGAACCCCG 4130
ATTCTGATCTCTCTCCGTTCTTGTACTTACCAAAACCAAAACCAACCTACTAATAAATATGATGAGACAA 4200
4210 4220 4230 4240 4250 4260 4270
TTGGAATTTGCTGCTCAATTTCT 4270
GTCTGCTCCCCCATAAAGACTCTTCCGCGGAAAAATGTTGCAACGGGAAGTGATATTCCGAGCAATTTCTG 4340
ACGTCGAGGCGCGAAAAACACATCTGCTGACAAAGAGTAAAGCAATTTCTCTCTCTCTCTCTCTCTCTCT 4410
TTCAATTCGTTTTTCAAAATGAGCTACTACAGAGTGAAAGAGCAAAATTTGCAAAACATTTTCTGTTGA 4480
GATGCACTTTTGAAAAAATTAACCTTTACGTTTCTAGATTTTATTTTTTATATAAATTAGAG 4550
4560 4570 4580 4590 4600 4610 4620
CTCTCTAGACCTGCTATATTTTAAAAACCT 4620
TCT 4690
TTCT 4760
ACAAGACTCGAAACGATGCTCAAAAGAGCAGTGAAGAAGAGTCCGGAATCTCTCTCTCTCTCTCTCTCT 4830
CCAACGTCATCATCGACGGAAGGTTCCCTAAGCATGCATTCCACATCTTCCAAGGTCTGTTGTTAGGAG 4900

Friday, 28 November 1997 12:03
pBS KS/X18

Page

4910 4920 4930 4940 4950 4960 4970
AACTGTTTTTTTTGTTTTCGTGACCTTCACATAGTCTCGGATGTTTATAAAAGTGAGGTCCTCGGACA 4970
CCTGCCATAAAATGTGAATCCGCCCAATGCTGGTACAAAAAACTTTGCAGAGCACCCTGCTTTATACATT 5040
TTTAGGATAAAATGTCATACGGGATTTGTCAAAACCCAAACTTTTTAAATTTTATTTTCAGATCAAAAATTG 5110
ATGTTAAAGTTTAAGATATTTACGAAAAAATGTTTTACTTAAACCTTTTATATGGATAAAATTTTAC 5180
AACATTCAGATAGGAGTTCGTCCTTAAACTTTTGTGGTCCGCCGAGAAGCTTTCGAATTAATAATCTCA 5250
6260 6270 6280 6290 6300 6310 6320
TTTTTAAGTTCGAAACTAATTTTTTGTGCAACTGAATTTTTCAGAGATGAACCTTAAAGTTTCAATTTAT 6320
CCCATTTGAAACCCTCCCTTCTATAAACTTCAAAATTTTCAGAGTTCAACGTCAGACGAAAAGTCTCC 6390
GTGATCAGACGATCTTACCTTAAGCGCTCCATCTGTGACAGCTATCAGACAGCCGATAGCCGCAACACCG 6460
GTCTCTCAAAATATTAACAAGCCTGTTGAGGTGAGTATTTTTGTTTCTGGGTAAGAGGCTTCTTGTCT 6530
AAGTTTGGCCTAAATTAATTAACCTTGGTTTAGAGGCTGGCAAGCCATTGATCAAGCATGGGCTAAACCT 6600
6610 6620 6630 6640 6650 6660 6670
GGGCGCGCTTGAACCATGTACAATTTCTTGGCCCGAGTAGTTGCAATCTAAAGATTCGAAGCTGGCTTC 6670
AAAGTCGGACTAGGCAAAAGTGCAAAAATGAAAAATGCTTTGAAATTCATACGCTTCCGCTCTTCCAT 6740
CTCTCTTTTTTGTGCTTTTTGTTGACAGATTTTCCCTTTTTTAGATTTTAAAGATTTTGATACACTTTAA 6810
TGCTGGCTTCGCTTCTTAAGAGCCTTCTATATTTTCGAAAAATAATCAATTTTTTAGGAAAAACCAAC 6880
ACTGGCAGTGAAAGGAGTGAAGAGCACAGCGAAAAAAGATCCACCTCCAGCTGTTCTCGCCACGTGACACC 6950
5960 5970 5980 5990 6000 6010 6020
CAGCCAACAATCGGAGTTGTTAGTCCATTATGGCACATAAGAAGTTGACAAATGGTACCTTTATTTCTG 6020
AACCTTACCTTATGTTTCGGTCCGTGACGTTTTTGTGACCATGTGATGGGAAGTAATTTTTGGATTATTT 6090
AAAGTTTGTGCGGGAATAGTAAGGAGGAGTACAATTATTTTATTTGTAGAAGGTTGAGTAAACTTTGAT 6160
TTTCTGACCATAAGTTTTTTTTCTGAAAGTTGTTTAAAAAATTCAGTTAAAAAATAAAAAATACCTCTA 6230
TAAAAATTTCTAAATTTCTGGGAATTTTTTTTCAAAAATGTTTTTCCAAAATAGTCTAATAGTAAGATTTG 6300
6310 6320 6330 6340 6350 6360 6370
TTTGTGAATTTACAAAACATATTTAAAAATACATTTTAAATTTATTTTCGATTTTTTTCGTTCCGAGGAACGAC 6370
AAAAAATCAGAAAAAGCGAATTTAATTTCCAAAAAATAATTTTGAAAAATCAGAAATTAAGCTACTTTT 6440
CAAAAAATCAACAAAAAATAATCAAAAGATTCTATTTTCAGAAATAGAGACATAATCAAAAAATATCA 6510
AAAAATCACTATTTTCCGGAATAAATTTGAGAAAAATTCAAAAATTTGAAAAATCAAAATTTGAGAAAA 6580
ATTTCCAGAAATTTAAAAAATAATTCCTTTGAGGAAAAATTTAAAAATTTTAAATGTGTATTTCTGAAACCA 6650
6660 6670 6680 6690 6700 6710 6720
AGCATTTTCCGACTTTTTCGGGATTTTCAGACTTGGGCTATAAATTTTGTCAAAAATAGGAAATCTAAA 6720
ATATTTGTATTTTTCGAAGCAATCTCTCTCAATCTCAAAATTCATATTTTATAATTTTCAGCCCGGTGATAT 6790
CTGAAAAACAGAACCTGAAAAAGCTCCAAATGAGCAATGACACGACGAGGCTTCCACCGCTTCCACC 6860
TCTAAAAATCAGTTGTTCCACTTAAATGACTTCAATCCGACAACCAACGTAAGATGTTCTCTTAAAA 6930
CAAGGAAAAATCACATCGCTGTCAAGTCGTTGGTCAAGTGCACCCCACTCCCAATTATATGACAAA 7000
7010 7020 7030 7040 7050 7060 7070
TGACCATTTTGCAGGATATGAGCAGTCTGTCGCTGTAAGACTCCATTTGTGGCTCTATGGCTCCGCTCAG 7070
GTGACTCCGCCGACAAAAACTTCTGGTAATCAATTCGCTGAGAGAGAAGGATGSGAAAATAAGACAATCAG 7140
GTAAATTTTGGAAACCTTGTATTTTCTGTTGAAAAATAGCTTCAAAATTAATAATTTAAAAAATCCGAG 7210
AAAAATGATGTTTGTCAAAAGGAAAACTTTTGATTTTTTGGTTTCTGAACTGTTCTGTTTAAAGTAACCA 7280
ACGTTGGAGCTCGTACCAAAAACTTTTCTTTTGATAATTTTGAATCTATAGTATTTTAAATTTTGGAA 7350

Friday, 28 November 1997 12:03
pBS KS/X18

Page

7360 7370 7380 7390 7400 7410 7420
TTGTGAAAG TTTCTTGAGATGTATTAAGT TTTAGGCATAGGCAGG TTTGTAGGCAGAAAGG TATGATG 7420
AGGCAGATAGGCTTGA TATTTACCAAGCCAAATAACAG TAAATAATATTTAAAAA AAACACTGAATAAA 7430
TCAAAAGCTAATAA TATTTGTTTATTGGACCTACCAACACCTTACATTTGCCCTACA TGTTCCTATATCC 7560
TTGTGTCTACATTTGAACGTTAATCACTAATTCGGTGAATGAACACTTGTAGAT TTTAATTTTCGACA 7630
GTAATTTTGGACACATTTGGCGTTAGAA TTTGAAAAAALACCTTCGACAGTTGAATCTTCATAACTCTCA 7700
7710 7720 7730 7740 7750 7760 7770
AAATA TTTCAAGATCCAGCGGCTACACCTCTGACGCCGG TTTTCCGATGTGCGCCAAATGAGGGAGAAG 7770
CTGAAAGAA TACGATGACATGACTCGTCGAGCACAGAACGGCTATCTGACAAG TGTGTTTGGTAGTAT 7840
TAGTTGTAGTCCCTTGACACACATATGAACACATTTCCGCTGCTCGTTTCGGTGGTCAAGGGAGCCATGGAGC 7910
AATTAATCCAGAAGGCTCAAAATTAATGAGCATCACTTGGTGA TCGAGGAATCCCCGAAAGAGCTTTGAT 7980
AGCATCTTCTTC TTTTGCATTCTTT TCTCTCTCTGCTGGGAG TCCCTGTTACACAGACATCTATTCAAG 8050
8060 8070 8080 8090 8100 8110 8120
CGCGAAGTGCAATTT TGGTTCTTAAAGATGAGAGGAGGAGGAGGAGCTATGAC TATGAATGAGCATCGA 8120
CGAACCGCGTGAAATAGTTTGGAGCTTAGATTGCAAATTACAGGAATATTTCCGG TGTACCTCAGTC TAC TT 8190
GGCATTTCGCGGCAGTTTGGTAAAC TTAGCCAAATTT TGTGTTGGGTCAAGCGTAATTTTCAAACAGTC 8260
GCAAGAATTTGCTCAACAAG TTTGCTGGCCCAAG TCTTTAGAGTC TGTACCAAGCTTGGTCCAATAC 8330
TTTTTGGGCCAAACTTTGGGAAGAAGTTGCCGCCAAATGCATTTTTCAAAAATAC TAGAACCACATCT 8400
8410 8420 8430 8440 8450 8460 8470
GGTTTAGCGCAAGTTGTCTCTAGGAGGATATCCAAATGT TATTTACTCTCTCTCT TTTCTAACGAAGAT 8470
CTATCGGTGCATCTCA TATGGAACCGATGCGGTTTTGCGCGCTTCGAAGAGCAT TCTTTGCTTT TGT T 8540
GCTTTTGTGTTATGCTTCTGAATTTGAAACTAGTTATGAAATTCAGATGTT TCAATGGAGATGCCAAG 8610
ATAGTTGGTATTACGAAAGTTTTCATGAATTTTGAATTTCTGTAGTTCTGTGAGTCT CATACATATAGCTC 8680
AAAAAGGTTCCGCTTGAGCCCTGCC TAAATCAAAATTTCTTTCAAGTCT TGTCCATTTAA TTAGGTT 8750
8760 8770 8780 8790 8800 8810 8820
ATGAAATAACGCGAACC TATCATTTCTCTTCGTTCTCAATTTCTTTTTCATGTACCC CCGATCC TGTGGGT 8820
TTCAATCTCTCTCAATTCACCTTAGATTGGCGCACAAAGCTCCCC TCTCTCTAATGCAAGCTGCTCTC 8890
CGGCTTCCATTT TGGCGACTTGC TCTTCCCTTCCCTTCCGTCGCCGTTTTCT TCCGATTGCC TGT 8960
GT TCCATATAATTAATTCATTAAGCCGAGAACAAACGAACGGGG TTTCTTTTTCTCTCTTACCATCTTT 9030
TGGCGGAATGAATTCGT TTTGTGATAGTGAAG TTTTGTCAAGAATTTTGG TTTTGGTAGCTTGGCA 9100
9110 9120 9130 9140 9150 9160 9170
AATATTGTTAGTATGCACCATGTTGCTCAGCATTTGTGTTCAA TATCTTGGTT TATTTCAAATTTGTT 9170
TCA TAAAAAC TTTGTAGGAGG TTTTCGTGGCGGACCCACTTTTCAAAC TCAATCCAAAAATGTTG 9240
TCAATTCGATCTCAAA TATCTCAACTATAACCTTAGAACGTTTTTT TCAAAATTTCCGAAATCAAT 9310
TTTTCGAAATTTCAAT TTTTAAAGAT TAACTGCCAAT TGAATTCATGCGCTTCCATTTGACTC 9380
CCATCATGAC TACGTTCCGTGGG TCTCGCCACAAAAA TACAGTTCC TCGGAAGGTT TTTGTGCGGGGC 9450
9460 9470 9480 9490 9500 9510 9520
CTAGCAGAGCCCATAGG TCCCAAAGCTCCCGATGAGCTGATAAAAATGTACTCTCAAGGAATATATGC 9520
AAAAATTTTGGAGCTTTT TACTAGAACAGTTTAAAGAAAGAAAAAGGTTTTTT TTAATTTAAAAA 9590
TTCAATTTTGAATATTAAAGCCAAATTTGAGTTGATCAC TTTGAGAAAAATTCAAAAT TGAAGGCTAG 9660
ATTTTGA TGAATTTATCTGGAAAGCGAACTTCTAAA TTAGAAAAAC TATAAAAAATCTAAATGTTT 9730
GAAATTTTTATT TGAATCTTAGTTGACTTTTT TGAATTTTCTAATTTGTTTCCAAGCTAATAGTCTTT 9800

Friday, 28 November 1997 12:03
pBS KS/X16

Page

9810 9820 9830 9840 9850 9860 9870
GAAGTCGCTCC TCAACCTCA AACCAGTGTGCCTCCATATTTGGAACACACACAAAGCAAAAACCAATIGA 9870
TACTATGTGTTTCGAGTAGCCACTTGACAAGAAGAAACTTGCCGACACITGGTGGCTGGTTCACCATTCCTCT 9940
CTCTTTGTCATTTGCAIAATCTTTCTCCCTCTTCTCATAAATAAATAAAGTGTGTCTCTGCGCTGCTCC 10010
CCGCTCTCGAGGGGGGGGGGGTACCCAGCTTTTGTTCCTCTTGTAGTGGGGTTAATTGGGCGCTTGGCGT 10080
AATCATGGTCATAGCTGTTTCTGTGTCAAATGTTTATCCGCTCACAATTCACACAACATACGAGCCGG 10150
10160 10170 10180 10190 10200 10210 10220
AAGCATAAAGTGTAAAGCTTGGGGTGGCTAATGAGTGAGCTAACTCACATTAATTGCGTGGCGCTCAGCTG 10220
CGCGCTTTCCAGTCGGGAACCTGTCTGTCAGCTGCATTAATGAATCGGCCAACGGCGGGGAGAGGGCG 10290
GTTTGGCTATTGGGGCGCTCTCCGCTTCTCTGCTCACTGACTCGCTGCGCTCGGTCTGTGGGCTGGGGCG 10360
AGCGGTATCAGCTCACTCAAAGCGGTAAACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC 10430
ATGTGAGCAAAAGGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCGCGCTTGTGGCGCTTTTCCATAGGC 10500
10510 10520 10530 10540 10550 10560 10570
TCGCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAATCCGACAGGACTATA 10570
AAGAATACAGSGCTTTCCCTTGGAAAGCTCCCTCGTGGCGCTCTCCGTTCGACCCGCGCGCTTACCGGA 10640
TACCCTGTCGCTTTCTCCCTTCCGGAAGCGTGGCGCTCTCTCATAGCTCAGCTGTAGGTATCTCAGCT 10710
CGGTGTAGGCTGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCGACCGCTGCGCTCT 10780
ATCCGCTAACTATCGCTCTGAGTCCAAACCGGTAAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGI 10850
10860 10870 10880 10890 10900 10910 10920
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCACAGAGTCTTGAAGTGGTGCCCAACGACGCT 10920
ACACTAGAAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCCTCGGA AAAAGAGTTGGTAG 10990
CTCTGTATCCGCAACAAACACCGCTGGTGGCGGTGGTCTTTTGTGTTGCAAGCAGCAGATTACCGCG 11060
AGAAAAAAGGATCTCAAGAAAGTCTTTGATCTTTTCTACGGGTCTGACGCTCACTGGAACGAAAC 11130
CAGCTTAAGGGATTTTGGTCAATGAGATATCAAAAAGGATCTTACCTAGATCTCTTAAATTA AAAATG 11200
11210 11220 11230 11240 11250 11260 11270
AAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTACCAATCTTAATCAGTGAG 11270
GCACCTATCTCAGCGATCTGTCTATTTCTGTCATCCATAGTTGCTTGACTCCCGCTGTGTAGATAAC 11340
CGATACGGGAGGGCTTACCATCTGGCGCCAGTGTGCAATGATACCGGAGACCCAAGCTCAGCGGCTCC 11410
AGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGTCTGCAACTTTATCCGCG 11480
TCCATCCAGCTATTAATTGTTCCCGGGAAGCTAGAGTAAGTAGTTCCGCGAGTTAAATAGTTTGGCGAAC 11550
11560 11570 11580 11590 11600 11610 11620
TTGTTGCCATTCCTACAGGCATCGTGGTGTACGCTGCTGCTTTGGTATGGCTTCACTCAGCTCCGGTTC 11620
CCACGATCAAGGCGAGTTACATGATCCCCATGTGTGCAAAAAGCGGTTAGCTCTCTGGTCTCCG 11690
ATCTTTGTGCAAGTAAGTTGGCGCGAGTGTATCACATCAAGTTATGGCAGCACTGCAAAATCTCTTA 11760
CTGTATGCCATCCGTAAGAATCTTTCTGTGACTGGTGAGTACTCAACCAAGTCACTTGAGAATAGTC 11830
TATGCGGCGACCGAGTTGCTCTTGGCGGGGCTCAATAGGGGATAATACCGCGCCACATAGCAGAACTTTA 11900
11910 11920 11930 11940 11950 11960 11970
AAAGTGCTCATCATTTGGAACAGTTCTTCTCGGGCGAAAACCTCAAGGATCTTACCGCTGTGAGATCCA 11970
GTTGCAATGTAACCCACTCGTGCACCAACATGATCTTCTCAGCATCTTTTACTTTCACCAAGCGTTCTGGG 12040
AGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGGAATAAGGGCGACACGGAAATGTGAACTACTATA 12110
CTCTTCTTTTTCAATATTAATGAAGCATTTATCAGGCTTATTGTCTCAATGAGCGGATACATATTTGAAT 12180
GATTTAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCGGAAAAGTCCCAT 12257

Tuesday, 18 November 1997 13:58

pLM3 (1 > 10847) Site and Sequence

Enzymes :

100 of 146 enzymes (Filtered)

Settings:

Linear, Certain Sites Only, Standard Genetic Code

(Seq 361)

```
GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTT
CTGCC TAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V
GGAGGTCGCTGAGTAGTGC CGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCGGAAC TGCGTGTAACTGACTTCTTAGACGAATCCCAATCCGCAAAACGC
G G R . V V R E Q N L S Y N K A R L D R Q L H E E S A . G . A F C
CTGCTTCGCGATGTACGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCCGTCTATATGCCCAACTGTAAC TAATAACTTGATCAATAATTATCATTAGT TAATGCCCGAGTAATCAAGTATCGGGTATAT
A A S R C T G O I Y A L T L I I D . L L I V I N Y G V I S S . P I Y
TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACGGGCGGACCGAC TGGCGGGTTGCTGGGGCGGGTAAC TGCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGACTATTTACGGTAAACTGCCCACTTGCCAGTACATCAAGTGATCATATGCCAAGTACGCC
TTGCGGTATCCCTGAAAGGTAAC TGCACTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCACTGATGTTACATAGTATACGGTTCATCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A
CCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCTGTAAGTGAATACCTGAAAGGATGAACCGTCACTGATGATGATCAATCAGT
P Y . R Q . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGCGTGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCCTACGCCAAAACCGTCACTGATGTTACCCGACCTATCGCCAACTGAGTGCCCTAAAGGTTACAGAGTGGGGTAAC TGCACTT
R Y Y H G D A V L A V H O W A V I A V . L T G I S K S P P H . R Q
TGGGAGTTTGTGTTGGCACAAATCAACGGGACTTTCCAAATGTCGTAACAAC TCCGCCCATTGACGCAAAATGGGCGGTAGGCGGTGACGGTGGGAG
ACCTCAAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTTGTTGAGGCGGGGTAAC TGCCTTTACCCGCCATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G
GTCTATATAAGCAGAGCTCTTGGCTAACTAGAGAACCCACTGCTTAC TGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGC
CAGATATATTCGCTCGAGAGACCGATTGATCTCTTGGGTGACGAATGACCGAATAGCTTTAATATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG
G L Y K Q S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A
TTTAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGTTATGGCTAGCATGACTGGTGGACAGCAAAATGGGTGGGGATCTGTACGAC
CAAAATTTGAATTCGAATGGTACCCCCCAAGAGTAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACTGTGCTTTACCCAGCCCTAGACATGCTG
F K L K L T M G G S H H H H H G M A S M T G G Q Q M G R D L Y D
```

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 2

GATGACGATAAGGTACCTAAGGATCCAGTGTGGTGAATTCTCGAGATATCGAATTCCTGCAGCCCCCTGCTCTTCAGCCAGATGCTGGACCCAGAGTCCC
CTACTGCTATTCCATGGATTCTAGGTCACACCACCTTAAGACGCTATAGCTTAAGGACGTCGGGGACGAGAAGTCGGTCTACGACCTGGGTCTCAGGG
100
D D D K V P K D P V V V N S A D I E F L O P L L F S O N L D P E S
AGAGAAAGAGGACAGTGCAGAATGTCCTGGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCACAGCTCCCTGGA
TCTCTTTCTCCTGTCACGTCTTACAGGACCTAGAGGCCGCTCTTGGACCTTCTCTGGTACAGGTCGGACGCTCCCAGGGTCCACTGAGTGTGAGGGGACC
120
insert pLM1
ORF pLM1
Q R K R T V Q N V L D L R Q N L E E T M S S L R G S Q V T H S S L E
GATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGTCCAGCCTCTCAACCGCTCGTCCCCCTGTGTCATGGCGCTATGGCCAGTCCAGTCCC
CTACTGGACGATGCTGTGCTACTACGGTTGGGTGCGTCGCACAGGTCGGAGAGGTTGGCGAGCAGGGGAGACAGTACC GCGATACCGGTCAGGTCAGGC
140
insert pLM1
ORF pLM1
M T C Y D S D D A N P R S V S S L S N R S S P L S V R Y G Q S S P
CGGCTGCAGGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCCGCTGGTACATGCACGGCGAACGGGCCACTACTCCCA
GCCGACGTCGACCACTGCGCGGGAGACCCACCCTCGACGGCGAGCCTCCCTGCGGGCGGACCATGTACGTGCCGCTTGCCCGGGTGATGAGGGTGT
160
insert pLM1
ORF pLM1
R L Q A G D A P S V G G S C R S E G T P A V Y M H G E R A H Y S H
CCATGCCCATGCGAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCTGAATCCCTGGACTCGGATGAGGTGGACCTCAAGTCCGGCTACA
GGTACGGGTACGCGTGGGGTCTTCGAGTCGGTATAGAGGGCGGACCTCGACCACTTAGGGACCTGAGCCTACTCCACCTGGASTTCAGGCCGATGT
180
insert pLM1
ORF pLM1
T M P M R S P S K L S H I S R L E L V E S L D S D E V D L K S G Y F
GAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCC
CTCGCTGCTACTGGAGTACCCGTTCTGGTACTGCCTCTACTACTGTAGTGATGGCCGACCTACTTTCGTGAGGTAGTCATCACTGAGTCGCTACGG
200
insert pLM1
ORF pLM1
S D S D L M G K T M T E D D D I T T G V D E S S S I S S G L S D A

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 3

TCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGCTTCTCGCAGGAAGTCAACAATAGTGTCTAC
AGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTTCATGAGGGTGACGAAGAGCGTCTTGAGTTGTTATCACGATG

insert pLM1

ORF pLM1

S D N L S S E E F N A S S S L N S L P S T P T A S R R N S T I V L

GCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTTAGTGAATCAGAGGAGAAAGCCCTAAAAAACTGGAGTACGACAGTGGTAG
CGTGTCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAAAATCAGTTAGTCTCCTCTTTCGGGGATTTCGACCTCATGCTGTCACCATC

insert pLM1

ORF pLM1

R T D S E K R S L A E S G L S V F S E S E E K A P K K L E Y D S G S

CCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTATCCAAAGGGTGAGAACTGAAAAAGCCCATCAGCCTG
GGACTTCTACCTTGGACCTGAAGATTACCGCCTCCCTCGCCGAGTCTCGACACTACTAAGTAGGTTCCCACTCTTGACTTTTTCGGGTAGTCGGAC

insert pLM1

ORF pLM1

L K M E P G T S K W R R E R P E S C D D S S K G G E L K K P I S L

GGCCACCCTGGTTCCTTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCATCACTCACACAGCCCAGAGTGCCCTCAAAGTCGCAGGCAAAAC
CCGGTGGGACCAAGGGACTTCTTCCCGTCTTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCGGGTCTCAGGGAGTTTCAGCGTCCGTTTG

insert pLM1

ORF pLM1

G H P G S L K K G K T P P V A V T S P I T H T A O S A L K V A G F

CTGAGGGCAAAGCTACAGACAAGGGTAAGCTTGCAGTGAAGAATACTGGGCTCCAACGCTCTCTCTGATGCTGGTGGGACCGCTGAGTGATGCTAA
GACTCCCGTTTCGATGCTGTTCCTTTCGAACGTCACCTCTTATGACCCGAGGTGCGAGGAGGAGACTACGACCAAGCCCTGGCGGACTCACTACGAT

insert pLM1

ORF pLM1

P E G K A T O K G K L A V K N T G L Q R S S S D A G R D R L S D A F

GAAGCCCCCTCGGGCATTGCTCGCCCTCCACTTCGGGATCCTTCGGCTACAAGAAGCCTCTCTGCCACAGGCACAGCCACTGTCATGCAAACTGGT
CTTCGGGGGAGCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAGCCGATGTTCTTCGGAGGAGGACGGTGTCGGTGTCGGTGACAGTACGTTTGACCA

insert pLM1

ORF pLM1

K P P S G I A R P S T S G S F G Y K K P P P A T G T A T V M Q T G

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GGTTCAGCCACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGACTAGCTTAGATGTTTCCAACAGCGCAGAGC
CCAAGTCGGTGAGAGTCGTCTAGGTCTTCAGGAGTCGGTAGGGACAGTTCCGGTCATTACCCGCGTTCTGATCGAATCTACAAAGGTTGTCGCGTCTCG 2300

insert pLM1

ORF pLM1

G S A T L S K I O K S S G I P V K P V N G R K T S L D V S N S A E
CAGGATTCCTGGCTCCGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCATGAGCGTGACCGGCGGGGCGGGTGG
GTCTTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTTCATGGCGTCGGACGGGGCCGGTCGGTTCAGTTCAAGATACTCGCACTGGCCGCGCCGCCCAAC 2400

insert pLM1

ORF pLM1

P G F L A P G A R S N I O Y R S L P R P A K S S S M S V T G G R G G
ACCTCGCCCTGTGAGCAGCAGCATTGACCCAGTCTCCTCAGCACCAAGCAGGGAGGCCTTACGCCCTTCAGACTGAAGGAGCCTACCAAGGTAGCCAGT
TGGAGCGGGACACTCGTCGCTGCTAACTGGGGTCAGAGGAGTCGTGGTTCGTCCCTCCGGAATGCGGAAGGTCTGACTTCCTCGGATGGTTCATCGGTCA 2500

insert pLM1

ORF pLM1

P R P V S S S I D P S L L S T K Q G G L T P S R L K E P T K V A S
GGGCGGACCACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGTGGCCTTGGACTCAGACAACATCTCCTTGAAGA
CCCGCCTGGTGAGGTGCGGGACAGTTAGTCTGCTTAGCCCTTTCTCTTCCGGTTTCGGTTCGGTCACCGGAACCTGAGTCTGTTGTAGAGGAACCTCT 2600

insert pLM1

ORF pLM1

G R T T P A P V N O T D R E K E K A K A K A V A L D S D N I S L)
GTATTGGCTCCCCAGAGAGTACTCCAAGAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAGGGCCACAGGGAA
CATAACCGAGGGGTCTCTCATGAGGGTCTTGGTTCGTTCCGTTGGGTGTCGGTGGTTCGACCGTCTCGACGGTGGTTGGGGAGAGTCCCGGTGTCCT 2700

insert pLM1

ORF pLM1

S I G S P E S T P K N Q A S H P T A T K L A E L P P T P L R A T A)
GAGCTTTGTCAAACCACTCTACTAGCCAATCTTGACAAGGTCAACTCCAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTG
CTCGAAACAGTTTGGTGGGAGTGATCGGTTAGAACGTTCAGTTGAGGTGTGACAGCTAGATGGTAGGTAGGTCACTATGGTGGGTACGAAGTTTCCAG 2800

insert pLM1

ORF pLM1

S F V K P P S L A N L D K V H S N S L D L P S S S D T T H A S I V

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 5

CCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCTGCTTACCCCCAGTCCGGCACCCATCCTCAATATTAATCAAGCCAGCTTCT
GGTCTAGACGTACGATGTTTCGAGTCGTAGACCCCCGGGAGAGGGAAGGACGAAGTGGGGTTCAGGCCGTGGGTAGGAGTTATAATTGAGTCGGTCAAGGA 2900

— insert pLM1 —

— ORF pLM1 —

P D L H A T S S A S G G P L P S C F T P S P A P I L N I N S A S F

CCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCAAACTCTCAGGCCTGCACAGGAGCATGGAGTCCCTCCAGAT
GGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCGGACGTGTCTCGTACCTCAGGGAGGTCTA 3000

— insert pLM1 —

— ORF pLM1 —

S Q G L E L M S G F S V P K E T R M Y P K L S G L H R S M E S L Q M

GCCAAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTGCTGCTCCACAGAAGAAGAGACGGAAGAGCTGACT
CGGTTACTCGGAGGGGTCACGAAGGGGTCGTATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAGGGTGTCTTCTTCTGCTTCTCGACTGA 3100

— insert pLM1 —

— ORF pLM1 —

P M S L P S A F P S S T P V P T P P A P P A A P T E E E T E E L T

TGGAGTGGGAAGCCCCAGAGCTGGGCAAC TGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGGGCTCAGGTACCAGCTTCAGTCCAGGAGG
ACCTCACCTTCGGGGTCTCGACCCGTTGACCTGTCATTAGTCGCCCTAGCCTTGTGAGAAGGGTCTTTCCCGAGTCCATGGTCGAAGTCAGGGTCCCTCC 3200

— insert pLM1 —

— ORF pLM1 —

V S G S P R A G Q L D S N Q R D R N T L P K K G L R Y Q L Q S Q E

AGACCAAGGAGAGGCGACATTCCCATACCATGGTGGGC TGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACCTTCCCATGCTCTGAG
TCTGGTTCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAAGAGGGGGACGTGAAGGGTACAGAGACTC 3300

— insert pLM1 —

— ORF pLM1 —

E T A E R R H S H T I G G L P E S D D Q S E L P S P P A L P M S L S

TGCCAAGGGCCAAC TTACCAACATAGTGAGTCCCCTGCGGCCACCAACGCCAAGAATCACCCGCTCCAACAGCATCCCCACCCACGAGGCGGCTTCGAG
ACGTTTCCCGGTTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGGGTTCTTAGTGGGCGAGGTTGTCTAGGGGTGGGTGCTCCGCCGGAAGCTC 3400

— insert pLM1 —

— ORF pLM1 —

A K G Q L T N I V S P T A A T T P R I T R S N S I P T H E A A F E

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 6

CTGTACAGCGGCTCCCAATGGGGAGCACCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTCAGGATCCTTCCGAGACCCACGGACGATGTT
GACATGTGCGGAGGGTTTACCCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCTTAGGAAGGCTCTGGGGTGCCGTGCTACAAAG 350

insert pLM1

ORF pLM1

L Y S G S Q M G S T L S L A E R P K G M I R S G S F R D P T D D V

ACGGCTCAGTGCTGTCCCTGGCCGCCAGTGCCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAATCCGGAAGCTTCGTAGGGAAC
TGCCGAGTCACGACAGGGACCGAGGTCACGGAGGAGGTGGATGAGGAGTCGACTCCTCTCCTACGTTAGACTCGTTTAGGCCCTCGAAGCATCCCTTGA 360

insert pLM1

ORF pLM1

H G S V L S L A S S A S S T Y S S A E E R M Q S E Q I R K L R R E L

GGAAATCATCCAGGAAAAAGTGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCC
CCTTAGTAGGGTCCCTTTTACCGGTGGAAGTGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAAAACCTCGTCTCGGACCACTTATACGTAGG 370

insert pLM1

ORF pLM1

E S S Q E K V A T L T S Q L S A N A N L V A A F E Q S L V N M T S

CGCCTGCGACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCATAGACTTTCTGAAGAAAAAGAACTCTGAGGCC
GCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCGTGACTCGACGACCTAAACGCTCTTTGGTATCTGAAAGACTTCTTTTCTTGAGACTCCGGS 380

insert pLM1

ORF pLM1

R L R H L A E T A E E K D T E L L D L R E T I D F L K K K N S E A

AUGCACTCATTGAGGGAGCCCTTAATGCCTCAGAAACACACCCAAAGAACTTCGGATCAAGAGACAAAACCTCTCAGATAGCATCTCAAGCCTCAACAG
TCCGTCASTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTTCCTGTTTGGAGGAGTCTATCGTAGAGTTCGGAGTTGTC 390

insert pLM1

ORF pLM1

Q A V I O G A L N A S E T T P K E L R I K R Q N S S D S I S S L N S

CATCACTAGCCATTCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAAGAAAAAAGAGTTGGGTCTATGAGCTTCGAAGTTCTTCAACAAA
GTAGTGATCGGTAAAGGTCGTAGCCGTCGTCGTTCCTACGACTACGCTTTTCTCTTTTCTCAACCCAGATACTCGAAGCTTCAAGGAAGTTGTT 400

insert pLM1

ORF pLM1

I T S H S S I G S S K D A D A K K K K K K S V V Y E L R S S F H I

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 7

```
GCGTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTCAGCCCCCTCATCCCCAACTAC
CGCAAGTCATATTTTTCCTCCGGGTTTCAGTCGAAGGAGTATGAGCCTATATCTCTCTAACGATGTGGGCTGAGAAGTCGGGGGAGTAGGGGTTGATG
insert pLM1
ORF pLM1
A F S I K K G P K S A S S Y S D I E E I A T P D S S A P S S P K L
AGCATGGTTCCACAGAGACTGCTTCACCTCCATCAAGTCCTCCACCTTGCTCCGTGGGCACTGATGTCACCGAGGGGCC TGCTCACCCAGCCCCCA
TCGTACCAAGGTGCTCTGACGAAGTGGGAGGTAGTTCCAGGAGTGGAACAGGAGGCACCCGTGACTACAGTGGCTCCCGGACGAGTGGGTCGGGGGGT
insert pLM1
ORF pLM1
Q H G S T E T A S P S I K S S T L S S V G T D V T E G P A H P A P H
CACTAGGCTGTTCCATGCAAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGAC
GTGATCCGACAAGGTACGTTACTCTCTCTCGGTCTCTTCTCTCCATAGCCTCGACGCGAGACTCGATACCCTCTCTCTTACTTCGAATGTCG
insert pLM1
ORF pLM1
T R L F H A N E E E E P E K K E V S E L R S E L V E K E M K L T D
ATCCGCTTGGAGGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAAATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATG
TAGGCGAACCTCCGGGAGTTGAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTTGTACGTCAACCTCCACCTGGACGACTTTCGTCCTTAC
insert pLM1
ORF pLM1
I R L E A L N S A H Q L D Q L R E T M H N M O L E V D L L K A E N
ACCGACTSAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCTTGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCTGGC
TGGCTGACTTCCATCGGGGTCCGGGAGTAGTCCGAGGTGAGGTCCGTCCTAGGGACCTAGTAGACGTAAATAGAAGGGGTGCGGCGAGGGATCCGGACC
insert pLM1
ORF pLM1
D R L K V A P G P S S G S T P G Q V P G S S A L S S P R R S L G L A
ACTCACCATTCCTTCGGCCCCAGTCTTGACAGACAGACCTGTCAACCATGGATGGCATCAGTACTTGTGGTCCAAGGAGGAAGTGACCTCCGGGTS
TGAGTGGGTAAGGAAGCCGGGTCAGAACGTCGTGTCTGGACAGTGGGTACCTACCGTAGTCATGAACACCAAGGTTTCTCTTCACTGGGAGGCCAC
insert pLM1
ORF pLM1
L T H S F G P S L A D T D L S P M D G I S T C G P K E E V T L R V
```

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

GTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCA3GAATTCTTCCTGGGCTGTAGCAAGGTCAGTGGAAAAGTTGACTGGGAAGA 4700
CAACCACTCC TACGGGGGCGTCGTGTAGTAGTTTCCCTGAAC TTCGTCGTCCTTAAGAAGGACCCGACATCGT TCCAGTCACCTTTTCAACTGACCTTCT

insert pLM1

ORF pLM1

V V R H P P Q H I I K G D L K Q Q E F F L G C S K V S G K V D V K

TGCTGGATGAAGCTGTTTTCAAGTGTTCAGGACTATATTTCTAAATGGACCCAGCCTCTACCCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAG 4800
ACGACCTACTTCGACAAAAGGTTCACAAGTTCCTGATATAAAGATTTTACCTGGGTCGGAGATGGGACCCCTGATTCGTGACTCAGGTAGGTACCGATGTC

insert pLM1

ORF pLM1

M L D E A V F Q V F K D Y I S K M D P A S T L G L S T E S I H G Y S

CATCAGCCAGTGAAACGAGTGTGGATGCAGAGCCCCCGAGATGCC TCTTCCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCCTGAAG 4900
GTAGTCGGTGCACTTTGCTCACAACCTACGTC TCGGGGGGCTCTACGGAGGAACGGCAGC TCCACAGTTATTGTATAGTCAGAGGGAGTTTCCAGACTTC

insert pLM1

ORF pLM1

I S H V K R V L D A E P P E M P P C R R G V N N I S V S L K G L I

GAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACC6GCGCTCGTCTCT 5000
CTCTTTACGCAGCTGTGCGACCACAAGCTCTGCGACTAGGGGTTGCGCTACTACGTCGTGATGTATTGCGAGGACGACTTCGTGGCCCGCGAGCAGSAGA

insert pLM1

ORF pLM1

E K C V D S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L

CGGGCCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGSAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCA 5100
GCCC6GGGTCGCGGTGCGCGTCTG6ATGGACTGGTTAGCGAACC6GCTCATGGACCACCTCGCGAGACCGGCACCTCCAGTGTCTCCCGTAGCAGTCTG

insert pLM1

ORF pLM1

S G P S G T G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T

CTTCAACATGCACCAAGCAGTCTTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCCAACCAGATAGACCGGAAACAGGAATTGGGGATGTGCCCTG 5200
GAAGTTGTACGTGGTCGTCAGAAGCTTCCTAGACGTTGACATAGAAAGGTTGATCGGTTGGTCTATCTGGCCCTTTGTCCTTAACCCCTACACGGGGAC

insert pLM1

ORF pLM1

F N M H Q Q S C K D L O L Y L S N L A N Q I D R E T G I G D V F L

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 4

```
GTGATTCTATTGGATGACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGCCCTATATTATAAGTA
CAC TAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCGGGAGTGGACGTTTCATAGTATTTACAGGGATATAATATCCAT
5300
-----insert pLM1-----
-----ORF pLM1-----
V I L L D D L S E A G S I S E L V N G A L T C K Y H K C P Y I I G
CCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCCT
GGTGGTTAGTCGGACATTTTACTGTGGGTTGGTACCGAAGCTGAAGTCTCAAC TGAAGAGGTTGTTGCACCTCGGTCGGTTACCGAAGGA
5400
-----insert pLM1-----
-----ORF pLM1-----
T T N Q P V K M T P N H G L H L S F R M L T F S N N V E P A N G F L
GGTTCGTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGG
CCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCTGTAGTTACGGTGTTCCTTCGACGAAGCCACGAGCTGACCCATGGGTTGCACACC
5500
-----insert pLM1-----
-----ORF pLM1-----
V R Y L R R K L V E S D S D I N A N K E E L L R V L D V V P K L V
TATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTCTGTCGTGCCATTGGCATTGAGGACTTCGGGA
ATAGTAGAGGTGTGGAAGGAACCTTCGTGTCTGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCT
5600
-----insert pLM1-----
-----ORF pLM1-----
Y H L H T F L E K H S T S D F L I G P C F F L S C P I G I E D F R
CCTGGTTTCATTGACCTGTGGAACAACCTATCATTCCTTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGSGAGGA
GGACCAAGTAAC TGGACACCTTGTTGAGATAGTAAGGGATAGATGTCTTCCTCGGTTCTACCTATTTCAGGTACCTGTCTTCGACGAACCTCCT
5700
-----insert pLM1-----
-----ORF pLM1-----
T V F I D L V N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D
CCAGTGGAATGGGTCGGGACACACTTCCTG3CCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGC
GGGTACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCS
5800
-----insert pLM1-----
-----ORF pLM1-----
P V E V V R D T L P V P S A Q Q D Q S K L Y H L P P P T V G P H S
```


Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 10

ATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTE 530X
TAACGGAGTGGAGGGCTCCATCCTGTCACTTCTGTCTGGGGTTCAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACITTGAACTCTTCGAC 530X
insert pLM1
ORF pLM1
I A S P P E D R T V K D S T P S S L D S D P L M A M L L K L Q E A
CCAACACATGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTACCCCCGGACAGCAGA 530X
GGTTGATGTAATCAGAGGTCTAGCTCTTTGGTAGGACCTGGGGTGGAAAGTCCGTGTGAAATCCCAAGCCGTTAGTGACAGTGGGGGCTGTCTGCT 530X
insert pLM1
ORF pLM1
A N Y I E S P D R E T I L D P N L Q A T L . G F G N H C H P R T A E
ACGCTGGCATCAGCTATCTTAGCTCCTCTCTCCCTCTCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGAGATGAA 530X
TGCGACCGTAGTCGATAGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGCGGGTCTCTCTTGTCTCTCTCTCTCTACTT 530X
insert pLM1
R V H Q L S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E
AGAGGAGGGACAGGTCTTGGTGCTGTACCTTTGAGAACCTCCTAGGAAGGAATGGTGGGTGGCGTTTGGGAACCTGTGCCCCCTAAACACATTTACTG 530X
TCTCTCTCTGTCCAAGAACCACGACATGGAACCTCTGAAGGATCCTTCTTACCACCCACCACCAACCTTGAACACGGGGGATTGTGTAAATGAC 530X
insert pLM1
R G G T G S V C C T F E N F L G R N G G V A F G N L C P L N T F T
GCCCTCTAATGACTTTGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTCATTACAAACCTCTGGGCTTTCTGGGGAGGAGGTTACAGAA 530X
CGGAGGAGATTACTGAAACCCCTTTCTACTAAGACCCAGAAAGGGAATGAAGAACAAGTAAATGTTTGAGGACCCGAAAGACCCCTCCCAAGCTT 530X
insert pLM1
G L L . . L V G K D D S G S F P . L L V S I T N S W A F V G G V Q
AACATCAAAACACTGCAGCAGTTCCTCCGGAATTCAGCTTGGACTTAACCAGGCTGAACCTTGC TCAAAAGAAGCCGAATTCAGCACACTGGCGCCGTTA 530X
TTGTAGTTTTGTGACGTCGTAAGGGGCTTAAGTCGAACCTGAATTGGTCCGACTTGAACGAGTTTCTTCGGCTTAAGGTCGTGTGACCGCCGGCAA 530X
insert pLM1
T S K H C S S S P E F S L D L T R L N L L K R S R I P A H W R P L
CTAGTTCTAGAGGGCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCCCCCCCGTGCTTCTTTGAC 530X
GATCAAGATCTCCCGGCAAAATTTGGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAACGGGGAGGGGGACGGAAGGAAGT 530X
L V L E G P F K P A D Q P R L C L L V A S H L L F A P P P C L P .
CCTGGAAGGTGCCACTCCCACTGTCTTTCTAATAAAATGAGGAAATGCATCGCATTTGCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGTTGGGG 530X
GGACCTTCCACGGTGAGGGTGACAGGAAAGGATTATTTACTCTTTAAGTAGCGTAACAGACTCATCCAGTAAGATAAGACCCCCACCCACCC 530X
P V K V P L P L S F P N K M R K L H R I V . V G V I L F V G V S V G

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 11

CAGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTC TGAGGCGGAAAGAACCAAGCTGGGGCTCTA
GTCTGTGCTTCCCCCTCTAACCCTTCGTTATCGTCCGTACGACCCCTACGCCACCGAGATACCGAAGACTCCGCTTTCTTGGTCGACCCCGAGAT 6700
R T A R G R I G K T I A G M L G M R V A L V L L R R K E P A G A L

GGGGGTATCCCCACGCGCCCTGTAGCGGCGCAATTAAGCGCGGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGC
CCCCATAGGGGTGCGCGGGACATCGCGCGTAATTCGCGCGCCACACCACCAATGCGCGTCGCACTGGCGATGTGAACGGTCGCGGGATCGCGGGCG 6800
G G I P T R P V A A H . A R R V V W L R A A . P L H L P A P . R P

TCCTTTCGCTTTCTCCCTTCCTTTCGCGCACGTTGCGCGGCTTCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTAGTGCTTTA
AGGAAAGCGAAAGAAAGGGAAGAGCGGTGCAAGCGCGCGAAAGGGGCAGTTGAGATTAGCCCCGTAGGGAATCCCAAGGCTAAATCACGAAAT 6900
L L S L S S L P F S P R S P A F P V K L . I G A S L . G S D L V L Y

CGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGT
GCCGTGGAGCTGGGGTTTTTGAAC TAATCCCACTACCAAGTGCATCACCCGGTAGCGGGACTATCTGCCAAAAGCGGGAAACTGCAACCTCAGGTGCA 7000
G T S T P K N L I R V M V H V V G H R P D R R F F A L . R V S P R

TCTTAAATAGTGGACTCTTGTTCAAACTGGAACAACAC TCAACCTATCTCGGTCTATTCTTTTGAATTATAAGGGATTTTGGGGATTTCGGCCTATTG
AGAAATTATCACCTGAGAACAGGTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAAACTAAATATCCCTAAAACCCCTAAAGCCGGATAAC 7100
S L I V D S C S K L E Q H S T L S R S I L L I Y K G F V G F R P I

GTTAAAAATGAGCTGATTAAACAAAAATTAACGCGAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCAAGGCTCCCAAGGCGAGC
CAATTTTTTACTCGACTAAATTGTTTTAAATTCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTTCAGGGGTCCGAGGGGTCCGTCGC
G . K M S . F N K N L T R I N S V E C V S V R V W K V P R L P R Q A 7200

AGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCAAGGCTCCCAAGCAGGCAGAAAGTATGCAAGCATGCATCTCAATT
TCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTCCACACCTTTTCAGGGGTCCGAGGGGTGTCGCTCTTCATACGTTTCGTACGTAGAGTTAA 7300
E V C K A C I S I S Q O P G V E S P Q A P Q Q A E V C K A C I S I

AGTCAGCAACCATAGTCCCGCCCTAAC TCCGCCATCCCGCCCTAAC TCCGCCAGTTCCGCCCATTCGCCCCCATGGCTGACTAATTTTTTTTAT
TCAGTCGTTGGTATCAGGGCGGGGATTGAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACC GACTGATTAAAAAAATA 7400
S O O P . S R P . L R P S R P . L R P V P P I L R P M A D . F F L

TTATGCAGAGGCGGAGGCGCCTCTGCC TCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTTAGGCTTTTGCAAAAAGCTCCCGGAGCT
AATACGTCTCCGGCTCCGGCGGAGACGGAGACTCGATAAGGTCTTCATCACCTCCGAAAAAACCTCCGGATCCGAAAACGTTTTTCGAGGGCCCTCGA 7500
F M Q R P R P P L P L S Y S R S S E E A F L E A . A F A K S S R E L

TGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTG
ACATATAGGTAAAGCTAGACTAGTTCTCTGTCTTACTCTAGCAAGCGTACTAATCTGTCTACCTAACGTGCGTCCAAAGAGGCGGGCAACCCACC 7600
V Y P F S D L I K R Q D E D R F A . L N K M O C T O V L R P L G W

AGAGGCTATTCGGCTATGACTGGGCAACAAGACAATCGGCTGCTCTGATGCCGCGTGTTCGGCTGTGACGCGAGGGGCGCCGGTCTTTTTGTCAA
TCTCCGATAAGCCGATAGTACCGTGTGTCTGTAGCGACGAGACTACGGCGGCAAGGCCGACAGTCGCGTCCCGCGGGGCAAGAAAAACAGTT 7700
R G Y S A M T G H N R O S A A L M P P C S G C Q R R G A R F F L S

GACCGACCTGTCCGGTCCCTGAATGAAGTGCAGGACGAGGACGCGCGCTATCGTGGCTGGCCACGACGCGCGTTCCTTGCGCAGCTGTGCTCGACGTT
CTGGCTGGACAGGCCACGGGACTTACTTGACGTCTGCTCCGTCGCGCGGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAA 7800
R P T C P V P . M N C R T R O R G Y R G V P R R A F L A Q L C S T L

Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 12

GTCAC TGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGG
CAGTGACTTCGCCCTTCCTGACCGACGATAACCCGCTTCACGGCCCCGCTCCTAGAGGACAGTAGAGTGGAACGAGGACGGCTCTTTCATAGGTAGTACC 7900
S L K R E G T G C Y V A K C R G R I S C H L T L L L P R K Y P S V

CTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCAGTACTCGGATGGAAAGC
GACTACGTTACGCCCGCAGCTATGCGAACTAGGCCGATGGACGGGTAAGCTGGTGGTTCGCTTGTAGCGTAGCTCGCTCGTGCATGAGCCTACCTTCG 8000
L M Q C G G C I R L I R L P A H S T Y K R N I A S S E H V L G W K

CGGTCTTGTGCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACGTGTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGAT
GCCAGAACAGCTAGTCTACTAGACCTGCTTCTCGTAGTCCCCGAGCGGGTCGGCTTGACAAGCGGTCCGAGTTCGCGCGTACGGGCTGCCGCTCTTA 8100
P V L S I R M I W T K S I R G S R Q P N C S P G S R R A C P T A R I

CTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAAAGGCCGCTTTTCTGGATTTCGACTGTGGCGGGCTGGGTGTGGCGGACC
GAGCAGCACTGGGTACCGCTACGGACGAACGGCTTATAGTACCACCTTTTACCGCGGAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCTGG 8200
S S . P M A M P A C R I S W V K M A A F L D S S T V A G W V V R T

GCTATCAGGACATAGCGTTGGCTACCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTACGGTATCGCCGCTCCCGA
CGATAGTCTGTATCGCAACCGATGGGCACTATAACGACTTCTCGAACC CGCGCTTACCGGACTGGCGAAGGAGCAGCAAATGCCATAGCGGCGAGGGCT 8300
A I R T . R W L P V I L L K S L A A N G L T A S S C F T V S P L P

TTGCGAGCGCATCGCTTCTATCGCTTCTTGACGAGTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAC
AAGCGTCGCTAGCGGAAGATAGCGGAAGAAGTGTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCGCTCGGGTTGGACGGTAGTG 8400
I R S A S P S I A F L T S S S E R D S G V R N D R P S D A Q P A I T

GAGATTTGATTCACCGCGCGCTTCTATGAAAGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCT
CTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCGGACCTACTAGGAGGTGCGGCCCTTAGAGTACGA 8500
R F R F H R R L L . K V G L R N R F P G R R L D O P P A R G S H A

GGAGTTCTTCCGCCACCCCAACTGTTTATTGCAGCTTATAATGGTTACAATAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCAGTS
CCTCAAGAAGCGGGTGGGGTTGAACAAATAACGTGCAATATTACCAATGTTTATTTCGTTATCGTAGTGTTTAAAGTGTTTATTCGTAAGAAAAAGTGAC 8600
S V L R P P O L V Y C S L . W L Q I K O . H H K F H K . S ! F F T

CATTCTAGTTGTGGTTTGTCCAAATCATCAATGTATCTTATCATGCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGT
GTAAGATCAACACCAACAGGTTTGAGTAGTTACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCAGCATTAGTACCAGTATCGACA 8700
A F . L W F V O T H Q C I L S C L Y T V D L . L E L G V I M V I A V

TTCTGTGTGAAATGTTATCCGCTCACAATCCACACAACATACGAGCGGGAAGCATAAAGTGTAAGCCTGGGGTGCTTAATGAGTGAGCTAATCTAC
AAGGACACACTTTAACAATAGGCGAGTGTTAAGGTGTGTGTATGCTCGGCCCTCGTATTTCACATTTCGGACCCACGGATTACTCACTCGATTGAGTG 8800
S C V K L L S A H N S T O H T S R K H K V . S L G C L M S E L T H

ATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGT
TAATTAACGCAACGCGAGTGACGGGCGAAAGGTCAGCCCTTTGGACAGCAGGTGACGTAATTACTTAGCCGGTTGCGCGCCCTCTCCGCCAAACGCA 8900
I N C V A L T A R F P V G K P V V P A A L M N R P T R G E R R F A

ATTGGGCGCTCTTCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTTGGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAACAGGTTAT
TAACCCGCGAGAAGGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTCGAGTGAGTTTCCGCCATTATGCCAATA 9000
V A L F R F L A H . L A A L G R S A A A S G I S S L K G G N T V I

uesday, 10 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 13

CCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTTCCATAG
GGTGTCTTAGTCCCTATTGCGTCTTTCTTGTACACTCGTTTTCCGGTCGTTTTCCGGTCTTGCCATTTTTCCGGCGCAACGACCGCAAAAAGGTATC 910
H R I R G . R R K E H V S K R P A K G Q E P . K G R V A G V F P .
GCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCTTGAAGC
CGAGCGGGGGGACTGCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACCGCTTTGGGCTGCTTGATTTCTATGGTCCGCAAGGGGGGACCTTCG 920
A P P P . R A S Q K S T L K S E V A K P D R T I K I P G V S P V K
TCCCTCGTGGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGTTTTCTCAATGCTACGCTGTA
AGGGAGCACGCGAGAGGACAAGGCTGGGACGGCAATGGCTATGGACAGCGGAAAGAGGGAAGCCCTTCGACCGCGAAAGAGTTACGAGTGCACAT 930
L P R A L S C S D P A A Y R I P V R L S P F G K R G A F S M L T L .
GGTATCTCAGTTCGGTGTAGGTGCTGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTAGCCCGACCGCTGCGCTTATCCGGTAACATATCGTCT
CCATAGAGTCAAGCCACATCCAGCAAGCGAGGTTGACCCGACACAGTGCTTGGGGGGCAAGTCGGGCTGGCGACCGGAATAGGCCATTGATAGCAGA 940
V S Q F G V G R S L Q A G L C A R T P R S A R P L R L I R . L S S
TGAGTCCAACCCGTAAGACACGACTTATGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTG
ACTCAGGTTGGGCCATTCTGTCTGAATAGCGGTGACCGTCTGCTGACCATGTCTTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAAC 950
V Q P G K T R L I A T G S S H V . Q D . Q S E V C R R C Y R V L
AAGTGGTGGCTAACTACGGCTACACTAGAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGAT
TTCACCACCGGATTGATGCCGATGTGATCTTCTGTCTATAAACCATAGACGCGAGACGACTTCGGTCAATGGAAGCCTTTTCTCAACCATCGAGAATA 960
E V V A . L R L H . K D S I V Y L R S A E A S Y L R K K S W . L L I
CCGGCAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTC
GGCGGTTTTGTTGGTGGCGACCATGCCACCAAAAAAACAAAGTTCGTCGCTAATGCGGCTTTTTTTCTAGAGTTCTCTAGGAACTAGAAAAAG 970
R O T N H R V . R V F F C L Q A A D Y A O K K R I S R R S F O L F
TACGGGGTCTGACGCTCAGTGAACGAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAA
ATGCCCCAGACTGCGAGTCACCTTGCTTTTGAAGTCAATTCCTAAACAGTACTCTAATAGTTTTCTAGAGTGGATCTAGGAAATTAATTTTT 980
Y G V . R S V E R K L T L R D F G H E I I K K D L H L D P F K L I .
TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
ACTTCAAAATTTAGTTAGATTTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCCGTGGATAGAGTCGCTAGACAGATAAAG 990
M K F . I N L K Y I . V N L V . O L P M L N Q . G T Y L S D L S I S
GTTATCCATAGTTGCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGGAGACCCACG
CAAGTAGGTATCAACGGACTGAGGGGACGACATCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGTCACGACGTTACTATGGCGCTCTGGGTGC 1000
F I H S C L T P R R V D N Y D T G G L T I W P Q C C N D T A R P T
CTCACCGGCTCCAGATTTATCAGCAATAAACGACGCGGGAAGGGCGGAGCGCAGAAGTGGTCTGCAACTTTATCCGCCCTCCATCCAGTCTATTAAT
GAGTGGCCGAGGTCTAAATAGTCGTTATTTGGTGGTCTGCGCTTCCCGGCTCGCGCTTACCAGGACGTTGAAATAGGCGGAGGTAGGTGAGATAATTA 1010
L T G S R F I S N K P A S R K G R A O K V S C N F I R L H P V Y .
TGTTCGCGGAAGCTAGAGTAAGTAGTTCCGCAAGTTAATAGTTTGGCAACGTTGTTGCCATTGCTACAGGCACTGTTGGTGTACGCTCGTCTTGGTA
ACAACGGGCTTCGATCTCATTCATCAAGCGGTCAATTAACAACGCGTTGCAACAACGGTAACGATGTCCGTAGCACACAGTGGGAGCAGCAACCAT 1020
L L P G S . S K . F A S . F A O R C C H C Y R H R G V T L V V V Y

• Tuesday, 18 November 1997 13:58
pLM3 (1 > 10847) Site and Sequence

Page 14

• TGGCTTCATTAGCTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTAGCTCCTTCGGTCTCCGATCGTTGT
ACCGAAGTAAGTCGAGGCCAAGGGTTGCTAGTTCGGCTCAATGTACTAGGGGGTACAACACGTTTTTTCGCCAATCGAGGAAGCCAGGAGGCTAGCAACA 1030
G F I O L R F P T I K A S Y M I P H V V Q K S G . L L R S S D R C
CAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GTCTTCATTCAACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTATTAAGAGAATGACAGTACGGTAGGCATTCTACGAAAAGACACTGACCA 1040
Q K . V G R S V I T H G Y G S T A . F S Y C H A I R K M L F C D V
GAGTACTCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAAGTT
CTCATGAGTTGGTTCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCGCAGTTATGCCCTATTATGGCGCGGTGTATCGTCTTGAA 1050
V L N Q V I L R I V Y A A T E L L L P G V N T G . Y R A T . Q N F
TAAAAGTGCTCATCATTTGAAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGACCCAA
ATTTTCACGAGTAGTAACCTTTTGCAAGAAGCCCCGCTTTTGAGAGTTCTTAGAATGGCGACAACCTTAGGTCAAGCTACATTGGGTGAGCACGTGGGTT 1060
K S A H H V K T F F G A K T L K D L T A V E I Q F D V T H S C T Q
CTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGT
GACTAGAAGTCGTAGAAAAATGAAGTGGTCGAAAGACCCACTCGTTTTTGTCTTCCGTTTTACGGCGTTTTTCCCTTATTCCCGCTGTGCCTTTACA 1070
L I F S I F Y F H Q R F V V S K N R K A K C R K K G N K G D T E M
TGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC 1080
ACTTATGAGTATGAGAAGGAAAAAGTTATAATAACTTCGTAATAGTCCCAATAACAGAGTACTCGCCTATGTATAAACTTACATAAATCTTTTATTG
L N T H T L P F S I L L K H L S G L L S H E R I H I . M Y L E K . T
AAATAGGGGTTCGCGCACATTTCCCGAAAAAGTGCCACCTGACGTC 10847
TTTATCCCAAGGCGCGTGTAAGGGGCTTTTCACGGTGGACTGCAG
N R G S A H I S P K S A T . R

Sequence of pTB72, an expression vector incorporating *C. elegans* UNC-53. The Open reading frame (ORF) of the prolinker + *Ce*-UNC-53 and (upper ORF) *Ce*-UNC-53 alone (lower ORF) are listed under the sequence.

[illegible]

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75

TCCCGTGAATGGTTTCATTTCGATTGTGGAATGAGAACTTCATTCATATTTGGAACGGTGTGGCTAGAGATGGCAAA
AAAAACCTTCGGTCCGTCGCACTTCCTCTCGAGGATCCCAACGCATCGTCTCTAAAAATGGCCGCTGGTTCGATGG
TGAAACCCCGGAGAAATGTGCTCAAAACGCTTCTCAACTCCAAAGACCTCGTCCGCTCAGCTCGCCAACTATCCCGAC
AACACGTCAATCCCTCGAGTCGTTGATCCAAATTCGATGCTACCAAGCATCAGACCATCGACAACATTTGAACAG
AAGACTCTAACTTCTCTCGCTCTCCCGCCGTTCTCTTATCTGCTGATCCGGTACCTGATGATATCCCACTTTTCC
CCCTTTTCCCGCAATTTCCGAGAAGCTGTTCCTTTGTTCTGATGCTCCCGGGTGGCGACGCCGAAGCG
ATTTAAAAACCTTTTCTTCCGAAACATTTCCCATTTGCTCATTTAATAGTCAAAATGAAATAAACAGTGTATGTACTT
AAAAAATAAATAAATAAATCGAGGGGGGCCCTATTGTTAGTGTACACTAAATGGCTAGAGCTCGTGATCGAC
CCTCGACTGTGCCCTTCTAGTGTCCAGGCATCTGTTGTTTGGCCCTCCCGCGTGCCTTCTGACCTGGAAGG
GCCACTCCCACTGTCTTCTTCCATAATAAATGAGGAAATTCGATCGCATATGTCTAGATAGTGTCTGATTTCTTCTG
GGGGGTGGGGTGGCGGACGACAGAGGGGGAGGATTGGGAAGACAATAGCAGGCAATGTCTGGGGATGCGGT
GGGCTCATGTGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTTAGGGGGATTCGCCACGCGCCCTGTAGCGG
CGCATTAAGCGCGGGGTGTGGTGTACGCGCAGCTGACCGTACACTGCCAGCGCCTAGCGCCCTAGCGCCCG
TCCTTTCCGTTTCTTCCCTTCTTCTTCCGACGTTTCCCGGCTTCCCGCTCAAGCTCTAAATCGGGCGATCC
CTTTAGGTTCCGATTAGTGTCTTACGCGCACTCGACGCCAAAAAATGATTAGGTTGATGTTTACAGTAT
GGGCACTCGCCCTGATAGCAGGTTTTCGCCCTTTGACGTTTGGATGCCAGTCTCTTTAATAGTGACTCTTGT
CCAACTGGGAACAACTCAACCTTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGGGATTTCGGCCTA
TTGGTAAAAATGAGCTGATTTAACAAAAATTAACGCGAAATTAATCTGTGGAATGTGTGTCAGTTAGGTTGT
GGAAGTCCCAAGCTCCCGACGAGGAGAAGTATGCAAAAGCATGCATCTCAATTAGTCAGCAACCAAGGTTG
GGAAGTCCCGGAGCTCCCGACGAGGACAAGTATGCAAAAGCATGCATCTCAATTAGTCAGCAACCAATGATCCG
GCCCTAACTCCGCCATCCCGCCCTTAAGTCCGCCCAATTCGCCGCTTCCGCCCATGGCTGACTAATTT
TTTTTATTATGCAAGGGCGAGGCCGCTCTGCCCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAA
GGCTAGGCTTTTCAAAAAGCTCCCGGGAGCTGTATATCCATTTTCGATTTTCGATGTAGTCAAGAGACAGATGAGG
ATCGTTTCGATGATTGAACAAGATGATTGACGAGGTTCTCGCCGCGTGGGTGGAGAGCTATTTCGGC
TATGACTGGGCAACACAGACAATCGGCTGCTGTGATCGCCGCGTGTTCGGGCTGTACGCGACGAGGGCGGGCG
GTCTTTTTGTCAAGACCGACCTGTCCGGTCCCTGAATGAATTCGACGAGGACGAGGACGCGGCTATCGTGGC
TGGCCACGACGGGCGTTCCTTGGCGAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGCACTGGCTGCTATT
GGCGCAAGTCCGGGGAGGATCTCTCTGATCTACCTTGTCTCTCCGGAAGAAGTATCCATCATGGGTGAT
GCAATCGCGGGCGTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGGAAACATGCGATCGAGCG
GAGCACTGCTCGGATGGAAGCGCGTCTTGTCGATCAGGATGATCGGACGAAGAGCATCAGGGGCTCGCGC
CAGCGAACTGTCCGAGGCTCAAGGCGCGCATGCCGACGCGGAGGATCTCGCTGAGCCCATGGCGGATG
CCTGCTTGGCAATATCATGGTGGAAATGGCGCTTTTCTGATTATCGACTGTGGCGGGTGGGTGTGGC
GGACCGCTTACGAGCATAGCTGTGGCTACCCGCTGATATTGCTGAAGAGCTGGCGGCAATGGGCTGACCGT
TTCTGTGCTTATCGGTATCGCGCTCCCGATTCGACGCGCATGCGCTTCTATCGCTTCTTGACAGTGTCTT
CTGAGCGGGAATCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCAGGAGATTTCTGATTCCA
CGCCGCGCTTCTATGAAAGTTGGGCTTCGGAATCGTTTTTCCGGAGCGCGGCTGGATGATCTCCCTCCAGCGGG
GGATCTCATGTGGAGTCTTCGCCCAACCACTTGTTTATGCGAGCTTAATAGTTTCAAAATAAGCAATAG
CATCACAAAATTCACAAATAAGCAATTTTTTCACTGCAATCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTT
ATCATGCTGTATACCGTGCACCTCTGAGCTGAGGCTGGCGTAATCATGTGTCATGCTGTTCTGTGTGAATT
GTTATCGGCTCACAATTCCACACAACATACGAGCGCGGAAGCATAAAGTGTAAAGCTGGGGTGGCTAATGAGTG
AGCTAACTCACATTAATGGTGTGGCTGACCTGCCCGCTTCCAGTTCGGGAACCTGTGCTGGCCAGCTGCATTA
ATGAATCGGCCAACCGCGGGGAGACCGGGTGTGCGTATTGGGCGCTCTCGCGTCTCGGCTCACTGACTCG
CTCGCGCTCGGCTGTCTCGGCTCGCGGAGCGGATCAGCTCACTCAAAAGCGGTAATACGGTTTATCCACAGAAT
CAGGGGATAACCGAGGAAGAACATGTGAGCAAAAGCGCAGCAAAAGGCCAGCAACCTGAAAAAGCGCGGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCGCTGACGAGCATCAAAAAATCGAGCTCAAGTCAGAGGTGGGGA
AACCAGCAGGACTATAAGGATACAGGCGTTTCCCGCTGGAAGCTCCGCTGCGCTCTCTGTCTCCGACCTT
CGCGCTTACCGGATACGTTCGCCCTTTCTCCCTTGGGAAGCGTGGCGGCTTTCTCAATGCTCAGCGTGTAGG
TATCTAGTTCCGTTGAGTGTGCTCGCTCCAAGCTGGGCTGTGTGCAGCAACCCCGTTACGCGCGACCGCT
CGCGCTTACCGGTAACTATCGTCTGAGTCCAACCGGTGAGCAGCAGCTTATCGCCACTGGCAGCGCCACT
GGTAACAGGATTAGCAGAGCGAGGATGTAGCGGGTGTACAGAGTCTTGAAGTGGTGGCTTAATACGCGGT
ACATAGAAGGACAGTATTGGTATCTCGGCTGCTGCTGAAGCCAGTATACCTTCGGAATAAGAGTGGTAGCTCT
TGATCCGGCAACCAACCGCTGTGTAGCGGTGGTTTTTTTGTGCAAGCAGAGTATCCGCGCAAAAAA
AGGATCTCAAGAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAACTCAGTAAAGGGA
TTTTGCTATGAGATTATCAAAAAGGATCTTCCAGCTAGATCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAA
AGTATATAGTAAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACTTATCAGGATCTGTCTA
TTTGGTCTACCATAGTTGCTGACTCGACCTCGCTGTAGTAGTACATACAGTACGGGAGGGCTTACCTTGGCC
CAGTGTGCAATGATACCGCGAGACCCACGCTCAGGGCTCAGAGTATACAGTAATACGAATAACCGACCGCGGA
GGGCGGAGCGCAGAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTTGTTGCCGGGAAGCTAGA
TGAAGTATCTCGCCAGTAAATAGTTTGGCAGACGTTGTGCCATGTCTACAGGATCGTGGGTGTACCGCTGCTG
GTTGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTACATGATCCCCATGTTGTGTGAAAA
AAGCGGTTAGCTCTTCCGCTCCCGATCGTTGTGCAAGATTAAGTTGGCGGAGTGTATCATCATGTGTTATG
CGACAGCTCATAAATTTCTTACTGTATCGCATCCGTAAGTACTGTTTCTGACTGGTGAGTACTCAACCAAG
TCAITCTGAGAATAGTGTATCGCGGACCGGAGTTGCTCTTGGCCGGGCTCAATACGGGATAATACCGCGCCACA
TAGCAGAAGCTTTAAAGTGTCTCATCTTGGAAAAAGTTCTCGGGCGGAAACTCTCAAGGATCTTACCGCTGT
GAGTCCAGTTCGAATGTAACCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTACCAGCGTTTCTG
GTGAGCGAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTGAATACATCAT
CTTCTCTTTTCAATTAATGTAAGCATTTACAGGTTATTGTTCTCATGAGCGGATACATATTGAATGTATTGAT
AAAAATAAACAATAAGGGTTCGCGGCAATTTCCCGAAAAAGTGCACCTGACGTGACGATCGGAGTCCGGAGATC
CCGATCCCTTATGGTGCATCTGATGATCTGCTGTGATCGCGCATAGTTAAGCAGATGTCTGCTCCCTGT
TTGTGTGTGGAGTTCGCTGAGTAGTGCCGAGCAAAATTAAGCTACAAAGCAAGGCAAGGCTGACCGCAAT
GCATGAAGAATCTGCTTAGGGTTAGGCGTTTTCGGCTGCTTCGCGATGACGGGCAAGATACCGCTGTGACAT
TGATTATTGACTAGTTAATAGTAACTCAATACGGGCTCATTAGTTTCATAGCCCATATATGAGGTTCGCGGTA
CATAACTTACGGTAAATGGCCCGCTGCTGACCGCCCAACGACCCCGCCCATGACGTGCAATAGTACGATG
GTTCCTCATAGTAACGCCAATAGGAGCTTTCAATTGACGTCAATGGGTGGAGCTATTTCAGGTAACCTGCCCACT
GGCAGTCATCATAGTGTATCATATGCGAAGTACGCGCCCTATTGACGCTCAATGACGGTA

FIG. 2.

Met Thr Thr Ser Asn Val Glu Leu Ile Pro Ile
Tyr Thr Asp Trp 15

Ala Asn Arg His Leu Ser Lys Gly Ser Leu Ser
Lys Ser Ile Arg 30

Asp Ile Ser Asn Asp Phe Arg Asp Tyr Arg Leu
Val Ser Gln Leu 45

Ile Asn Val Ile Val Pro Ile Asn Glu Phe Ser
Pro Ala Phe Thr 60

Lys Arg Leu Ala Lys Ile Thr Ser Asn Leu Asp
Gly Leu Glu Thr 75

Cys Leu Asp Tyr Leu Lys Asn Leu Gly Leu Asp
Cys Ser Lys Leu 90

Thr Lys Thr Asp Ile Asp Ser Gly Asn Leu Gly
Ala Val Leu Gln 105

Leu Leu Phe Leu Leu Ser Thr Tyr Lys Gln Lys
Leu Arg Gln Leu 120

Lys Lys Asp Gln Lys Lys Leu Glu Gln Leu Pro
Thr Ser Ile Met 135

Pro Pro Ala Val Ser Lys Leu Pro Ser Pro Arg
Val Ala Thr Ser 150

Ala Thr Ala Ser Ala Thr Asn Pro Asn Ser Asn
Phe Pro Gln Met 165

Ser Thr Ser Arg Leu Gln Thr Pro Gln Ser Arg
Ile Ser Lys Ile 180

Asp Ser Ser Lys Ile Gly Ile Lys Pro Lys Thr
Ser Gly Leu Lys 195

Pro Pro Ser Ser Ser Thr Thr Ser Ser Asn Asn
Thr Asn Ser Phe 210

Arg Pro Ser Ser Arg Ser Ser Gly Asn Asn Asn
Val Gly Ser Thr 225

Ile Ser Thr Ser Ala Lys Ser Leu Glu Ser Ser
Ser Thr Tyr Ser 240

FIG. 2 CONTINUED.

Ser Ile Ser Asn Leu Asn Arg Pro Thr Ser Gln
Leu Gln Lys Pro 255

Ser Arg Pro Gln Thr Gln Leu Val Arg Val Ala
Thr Thr Thr Lys 270

Ile Gly Ser Ser Lys Leu Ala Ala Pro Lys Ala
Val Ser Thr Pro 285

Lys Leu Ala Ser Val Lys Thr Ile Gly Ala Lys
Gln Glu Pro Asp 300

Asn Ser Gly Gly Gly Gly Gly Gly Met Leu Lys
Leu Lys Leu Phe 315

Ser Ser Lys Asn Pro Ser Ser Ser Ser Asn Ser
Pro Gln Pro Thr 330

Arg Lys Ala Ala Ala Val Pro Gln Gln Gln Thr
Leu Ser Lys Ile 345

Ala Ala Pro Val Lys Ser Gly Leu Lys Pro Pro
Thr Ser Lys Leu 360

Gly Ser Ala Thr Ser Met Ser Lys Leu Cys Thr
Pro Lys Val Ser 375

Tyr Arg Lys Thr Asp Ala Pro Ile Ile Ser Gln
Gln Asp Ser Lys 390

Arg Cys Ser Lys Ser Ser Glu Glu Glu Ser Gly
Tyr Ala Gly Phe 405

Asn Ser Thr Ser Pro Thr Ser Ser Ser Thr Glu
Gly Ser Leu Ser 420

Met His Ser Thr Ser Ser Lys Ser Ser Thr Ser
Asp Glu Lys Ser 435

Pro Ser Ser Asp Asp Leu Thr Leu Asn Ala Ser
Ile Val Thr Ala 450

Ile Arg Gln Pro Ile Ala Ala Thr Pro Val Ser
Pro Asn Ile Ile 465

Asn Lys Pro Val Glu Glu Lys Pro Thr Leu Ala
Val Lys Gly Val 480

Lys Ser Thr Ala Lys Lys Asp Pro Pro Pro Ala
Val Pro Pro Arg 495

Asp Thr Gln Pro Thr Ile Gly Val Val Ser Pro
Ile Met Ala His 510

FIG. 2 CONTINUED.

Lys Lys Leu Thr Asn Asp Pro Val Ile Ser Glu
Lys Pro Glu Pro 525

Glu Lys Leu Gln Ser Met Ser Ile Asp Thr Thr
Asp Val Pro Pro 540

Leu Pro Pro Leu Lys Ser Val Val Pro Leu Lys
Met Thr Ser Ile 555

Arg Gln Pro Pro Thr Tyr Asp Val Leu Leu Lys
Gln Gly Lys Ile 570

Thr Ser Pro Val Lys Ser Phe Gly Tyr Glu Gln
Ser Ser Ala Ser 585

Glu Asp Ser Ile Val Ala His Ala Ser Ala Gln
Val Thr Pro Pro 600

Thr Lys Thr Ser Gly Asn His Ser Leu Glu Arg
Arg Met Gly Lys 615

Asn Lys Thr Ser Glu Ser Ser Gly Tyr Thr Ser
Asp Ala Gly Val 630

Ala Met Cys Ala Lys Met Arg Glu Lys Leu Lys
Glu Tyr Asp Asp 645

Met Thr Arg Arg Ala Gln Asn Gly Tyr Pro Asp
Asn Phe Glu Asp 660

Ser Ser Ser Leu Ser Ser Gly Ile Ser Asp Asn
Asn Glu Leu Asp 675

Asp Ile Ser Thr Asp Asp Leu Ser Gly Val Asp
Met Ala Thr Val 690

Ala Ser Lys His Ser Asp Tyr Ser His Phe Val
Arg His Pro Thr 705

Ser Ser Ser Ser Lys Pro Arg Val Pro Ser Arg
Ser Ser Thr Ser 720

Val Asp Ser Arg Ser Arg Ala Glu Gln Glu Asn
Val Tyr Lys Leu 735

Leu Ser Gln Cys Arg Thr Ser Gln Arg Gly Ala
Ala Ala Thr Ser 750

Thr Phe Gly Gln His Ser Leu Arg Ser Pro Gly
Tyr Ser Ser Tyr 765

Ser Pro His Leu Ser Val Ser Ala Asp Lys Asp
Thr Met Ser Met 780

FIG. 2 CONTINUED.

His Ser Gln Thr Ser Arg Arg Pro Ser Ser Gln
Lys Pro Ser Tyr 795

Ser Gly Gln Phe His Ser Leu Asp Arg Lys Cys
His Leu Gln Glu 810

Phe Thr Ser Thr Glu His Arg Met Ala Ala Leu
Leu Ser Pro Arg 825

Arg Val Pro Asn Ser Met Ser Lys Tyr Asp Ser
Ser Gly Ser Tyr 840

Ser Ala Arg Ser Arg Gly Gly Ser Ser Thr Gly
Ile Tyr Gly Glu 855

Thr Phe Gln Leu His Arg Leu Ser Asp Glu Lys
Ser Pro Ala His 870

Ser Ala Lys Ser Glu Met Gly Ser Gln Leu Ser
Leu Ala Ser Thr 885

Thr Ala Tyr Gly Ser Leu Asn Glu Lys Tyr Glu
His Ala Ile Arg 900

Asp Met Ala Arg Asp Leu Glu Cys Tyr Lys Asn
Thr Val Asp Ser 915

Leu Thr Lys Lys Gln Glu Asn Tyr Gly Ala Leu
Phe Asp Leu Phe 930

Glu Gln Lys Leu Arg Lys Leu Thr Gln His Ile
Asp Arg Ser Asn 945

Leu Lys Pro Glu Glu Ala Ile Arg Phe Arg Gln
Asp Ile Ala His 960

Leu Arg Asp Ile Ser Asn His Leu Ala Ser Asn
Ser Ala His Ala 975

Asn Glu Gly Ala Gly Glu Leu Leu Arg Gln Pro
Ser Leu Glu Ser 990

Val Ala Ser His Arg Ser Ser Met Ser Ser Ser
Ser Lys Ser Ser 1005

Lys Gln Glu Lys Ile Ser Leu Ser Ser Phe Gly
Lys Asn Lys Lys 1020

Ser Trp Ile Arg Ser Ser Leu Ser Lys Phe Thr
Lys Lys Lys Asn 1035

Lys Asn Tyr Asp Glu Ala His Met Pro Ser Ile
Ser Gly Ser Gln 1050

FIG. 2 CONTINUED.

Gly Thr Leu Asp Asn Ile Asp Val Ile Glu Leu
Lys Gln Glu Leu 1065

Lys Glu Arg Asp Ser Ala Leu Tyr Glu Val Arg
Leu Asp Asn Leu 1080

Asp Arg Ala Arg Glu Val Asp Val Leu Arg Glu
Thr Val Asn Lys 1095

Leu Lys Thr Glu Asn Lys Gln Leu Lys Lys Glu
Val Asp Lys Leu 1110

Thr Asn Gly Pro Ala Thr Arg Ala Ser Ser Arg
Ala Ser Ile Pro 1125

Val Ile Tyr Asp Asp Glu His Val Tyr Asp Ala
Ala Cys Ser Ser 1140

Thr Ser Ala Ser Gln Ser Ser Lys Arg Ser Ser
Gly Cys Asn Ser 1155

Ile Lys Val Thr Val Asn Val Asp Ile Ala Gly
Glu Ile Ser Ser 1170

Ile Val Asn Pro Asp Lys Glu Ile Ile Val Gly
Tyr Leu Ala Met 1185

Ser Thr Ser Gln Ser Cys Trp Lys Asp Ile Asp
Val Ser Ile Leu 1200

Gly Leu Phe Glu Val Tyr Leu Ser Arg Ile Asp
Val Glu His Gln 1215

Leu Gly Ile Asp Ala Arg Asp Ser Ile Leu Gly
Tyr Gln Ile Gly 1230

Glu Leu Arg Arg Val Ile Gly Asp Ser Thr Thr
Met Ile Thr Ser 1245

His Pro Thr Asp Ile Leu Thr Ser Ser Thr Thr
Ile Arg Met Phe 1260

Met His Gly Ala Ala Gln Ser Arg Val Asp Ser
Leu Val Leu Asp 1275

Met Leu Leu Pro Lys Gln Met Ile Leu Gln Leu
Val Lys Ser Ile 1290

Leu Thr Glu Arg Arg Leu Val Leu Ala Gly Ala
Thr Gly Ile Gly 1305

Lys Ser Lys Leu Ala Lys Thr Leu Ala Ala Tyr
Val Ser Ile Arg 1320

FIG. 2 CONTINUED.

Thr Asn Gln Ser Glu Asp Ser Ile Val Asn Ile
Ser Ile Pro Glu 1335

Asn Asn Lys Glu Glu Leu Leu Gln Val Glu Arg
Arg Leu Glu Lys 1350

Ile Leu Arg Ser Lys Glu Ser Cys Ile Val Ile
Leu Asp Asn Ile 1365

Pro Lys Asn Arg Ile Ala Phe Val Val Ser Val
Phe Ala Asn Val 1380

Pro Leu Gln Asn Asn Glu Gly Pro Phe Val Val
Cys Thr Val Asn 1395

Arg Tyr Gln Ile Pro Glu Leu Gln Ile His His
Asn Phe Lys Met 1410

Ser Val Met Ser Asn Arg Leu Glu Gly Phe Ile
Leu Arg Tyr Leu 1425

Arg Arg Arg Ala Val Glu Asp Glu Tyr Arg Leu
Thr Val Gln Met 1440

Pro Ser Glu Leu Phe Lys Ile Ile Asp Phe Phe
Pro Ile Ala Leu 1455

Gln Ala Val Asn Asn Phe Ile Glu Lys Thr Asn
Ser Val Asp Val 1470

Thr Val Gly Pro Arg Ala Cys Leu Asn Cys Pro
Leu Thr Val Asp 1485

Gly Ser Arg Glu Trp Phe Ile Arg Leu Trp Asn
Glu Asn Phe Ile 1500

Pro Tyr Leu Glu Arg Val Ala Arg Asp Gly Lys
Lys Thr Phe Gly 1515

Arg Cys Thr Ser Phe Glu Asp Pro Thr Asp Ile
Val Ser Lys Lys 1530

Trp Pro Trp Phe Asp Gly Glu Asn Pro Glu Asn
Val Leu Lys Arg 1545

Leu Gln Leu Gln Asp Leu Val Pro Ser Pro Ala
Asn Ser Ser Arg 1560

Gln His Phe Asn Pro Leu Glu Ser Leu Ile Gln
Leu His Ala Thr 1575

Lys His Gln Thr Ile Asp Asn Ile

FIG. 3. : tblastn search of the EST division of Genbank with the ORF of the longest known *Ce*-UNC-53 cDNA, tb3-M5, reveals two EST's with homology to a predicted coiled-coil region in *Ce*-UNC-53.

TBLASTN 1.4.9MP [26-March-1996] [Build 14:27:13 Apr 1996]

Reference: Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-10.

Query= tb3 M5 ORF

(1583 letters)

Database: Non-redundant Database of GenBank EST Division
647,253 sequences: 234,216,808 total letters.

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum P(N)	Probability N
dbj D35780 CELK025D6F C.elegans cDNA clone yk25d6 : 5'...	+2	358	7.9e-54	3
dbj D33048 CELK025D6R C.elegans cDNA clone yk25d6 : 3'...	-1	177	8.6e-16	1
gb H09036 H09036 y196cl1.r1 Homo sapiens cDNA clo...	+1	115	1.1e-05	1
gb AA049124 AA049124 mj46f04.c1 Soares mouse embryo N...	+3	106	8.6e-05	1
gb R91475 R91475 yq08cl1.c1 Homo sapiens cDNA clo...	+2	59	0.21	2
gb T23446 T23446 seq2955 Homo sapiens cDNA clone ...	-1	61	0.99	2
gb R86390 R86390 SW1ICA339SK Brugia malayi infect...	+2	74	0.996	1
gb T44781 T44781 8044 Arabidopsis thaliana cDNA c...	+1	71	0.9992	1
gb T75582 T75582 yd63f11.r1 Homo sapiens cDNA clo...	+2	64	0.99992	2

gb|H09036|H09036 y196cl1.r1 Homo sapiens cDNA clone 46037 5'.
Length = 489

Plus Strand HSPs:
Score = 115 (52.1 bits), Expect = 1.1e-05, P = 1.1e-05
Identities = 22/70 (31%), Positives = 45/70 (64%), Frame = +1

Query: 1059 IELKQELKERDSALYEVRLDNLDRAREVDVLRETIVNKLKTENKOLKKEVDKLTNGPATRA 1118
++L+ EL+... L ++RL+ L A ++D LRE +N++++E ++LK D+L +
Sbjct: 7 HOLRNLRLDKEMKLTDIRLEALSSAHOLOLREAHNRMOSEIEKLKXNDRLKSESOQSG 186

Query: 1119 SSRASIPVIY 1128
SR S P ++
Sbjct: 187 CSRGSPSVH 216

gb|AA049124|AA049124 mj46f04.c1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA clone 479167 5'.
Length = 337

Plus Strand HSPs:
Score = 106 (48.0 bits), Expect = 8.6e-05, P = 8.6e-05
Identities = 23/58 (39%), Positives = 38/58 (65%), Frame = +3

Query: 1057 DVIELKQELKERDSALYEVRLDNLDRAREVDVLRETIVNKLKTENKOLKKEVDKLTNGP 1114
+V EL+ EL E++ L ++RL+ L A ++D LAET++ ++ E LK E D+L P
Sbjct: 99 EVSELASELWEKEMKLTDIRLEALNSAHOLOLRETMMHNOLEVDLLKAENDRLKVP 272

FIG. 4.

A Search of the Genbank databases with part of the nucleotide binding domain of *Ce-UNC-53* does not identify statistically significant proteins except for the *C. elegans* cosmid containing *Ce-unc-53*.

```

5  TBLASTN 1.4.8MP [20-June-1995] [Build 18:00:05 Aug 29 1995]
   Query= section5 (240 letters)
   >1clisectio5
   ILTERRVLVLAGATGIGKSKLAKTLAATVVSIRTNOSEDSIVNISIPENNKELLQVERRLE
10  KILRSKESCIVILDNIPKNRIAFVVSFANVPLONNEGFVVTNRYOIPLOIHNNFK
   MSVMSNRLEGFILRYLRRRAVEDEYRLTVQMPSELFKIIDFFPIALQAVNNFIKTNVSVO
   VTVGPRACLNCLTVGGSREWFIRLWNNFIPYLERVAROGKKNLRLSLHFLRGSHRHRLX

   Database: Non-redundant PDB+GBupdate+GenBank+EMBLupdate+EMBL
             520,383 sequences; 367,017,413 total letters.

15                                     Smallest
                                     Sum
                                     Probability
Sequences producing High-scoring Segment Pairs:
emb1247810|CEF45E10 Caenorhabditis elegans cosmid F45... -2 1131 5.1e-158 2
gb1R41071|R41071 Hk575-f Homo sapiens cDNA clone k... +2 53 0.33 2
20  gb1T44781|T44781 8044 Arabidopsis thaliana cDNA cl... +1 74 0.35 1
   emb1248334|CEF10B5 Caenorhabditis elegans cosmid F10... +3 71 0.83 3
   gb1M81884|EPFCPCG Epifagus virginiana chloroplast c... +1 49 0.91 4
   gb1L09547|PEAPCLP Pisum sativum (clone pCLp) nuclea... +1 71 0.99 1
25  gb1M32604|TOMCD4B Tomato ATP-dependent protease (CD... +1 71 0.99 1
   emb1X69188|APTUSGA A.pyhilitidis mRNA for gamma-tubulin +2 56 0.992 3
   gb1T44782|T44782 8045 Arabidopsis thaliana cDNA cl... +1 68 0.9995 1
   gb1M17087|HUMRASK12 Human c-ras-K1-2 activated oncoge... +2 58 0.9998 1
   emb1X57702|IGGNATRIIP G.gallus RNA for precursor of nat... +3 56 0.9999 2
30  gb1K015201|HUMRASKB1 Human lung adenocarcinoma (PR371)... +2 57 0.99995 1

>gb1R41071|R41071 Hk575-f Homo sapiens cDNA clone k575-f.
   Length = 310
   Plus Strand HSPs:

35  Score = 53 (24.5 bits), Expect = 0.40, Sum P(2) = 0.33
   Identities = 9/15 (60%), Positives = 13/15 (86%), Frame = +2

Query: 130 GFILRYLRRRAVEDE 144
      GF+RYLRR+ VE +
40  Sbjct: 26 GFLVRYLRRLVESD 70

   Score = 47 (22.7 bits), Expect = 0.40, Sum P(2) = 0.33
   Identities = 9/26 (34%), Positives = 17/26 (65%), Frame = +3

45  Query: 170 NNFIEKTNVSVDVTGPRACLNCLTV 195
      F-EK ---O -GP L+ PL +
   Sbjct: 147 HTFLEKXSTLDFLIGPCFFLSGPLAL 224

```

FIG. 5.

Three frame translation of EST gb:R41071.

Regions of homology region with *Ce-Unc-53* in two different frames are underlined. The spacing between the blocks of homology is of similar size as that in *Ce-UNC-53*.

- 5 Subsequent re-cloning and re-sequencing of this region in man identified multiple sequencing errors gb:R41071, and identified an ORF which is more homologous to and co-linear with *Ce-UNC-53* (see alignment in fig 12).

10
 15
 20
 25
 30
 35
 40
 45

```

CTCCAACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGT
  10      20      30      40      50      60
  L Q Q R G A S Q W L P G S L P E E E A G
  S N N V E P A N G F L V R Y L R R K L V
  P T T W S Q P M A S W F V T * G G S W *

AGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGGTGCTCGACTTGG
  70      80      90      100     110     120
  R V R Q R H Q C Q Q G R A A S G C S T W
  E S D S D I N A N K E E L L R G A R L G
  S Q T A T S M P T R K S C F G V L D L G

GTACCCAAGCCTGTGGTATCATCTTCCACACCTTCCTTGAGAAGCACAGCACCTTAGACT
  130     140     150     160     170     180
  V P K P V V S S S T P S L R S T A P * T
  Y P S L W Y H L P H L P * E A Q H L R L
  T Q A C G I I F H T F L E K H S T L D F

TTCTCATCGGCCCTTGCTTCTTTCTGTGCGGTCCATTGGCATTGAGGCTTCCGGACCTTG
  190     200     210     220     230     240
  F S S A L A S F C R V H W H * G F R T L
  S H R P L L L S V G S I G I E A S G P C
  L I G P C F F L S G P L A L R L P D L V

TTTATTGACCTGTGGACAACCTATCATTTCTATCTACAGGAGGAGCCAAGGATTGGAT
  250     260     270     280     290     300
  F I D L W T T L S F P I Y R R S Q G L D
  L L T C G Q L Y H F L S T G G A K D W I
  Y * P V D N S I I S Y L Q E E P R I G *

AAAGGTCCAT
  310
  K G P
  K V H
  R S
  
```


FIG. 6 : blastn search of the EST division of Genbank with Hu-unc-53/1 cDNA 3b.

BLASTN 1.4.9MP (26-March-1996) (Build 14:27:07 Apr 1 1996)					
Query= Hu-unc-53/1 cDNA 3b					
(3256 letters)					
Database: Non-redundant Database of GenBank EST Division					
647,253 sequences; 234,216,808 total letters.					
Sequences producing High-scoring Segment Pairs:					
		High	Smallest		
		Score	Sum	Probability	LOCUS
			P(N)	N	assignment
53/1	yx91b09.r1 Homo sapiens cDNA clone 2...	1668	2.1e-130	1	hu-UNC-
25	qb1AA043997/AA043997 zk58a01.r1 Soares pregnant uterus Nb...	1316	8.3e-129	3	hu-UNC-53/1
	qb1AA049124/AA049124 mj46204.r1 Soares mouse embryo NbME1...	1324	9.1e-102	1	ms-UNC-53/1
	qb1T05560/T05560 ESTC3449 Homo sapiens cDNA clone HFB...	892	5.1e-84	3	hu-UNC-53/1
	qb1N24681/N24681 yx91b09.s1 Homo sapiens cDNA clone 2...	782	9.9e-75	2	hu-UNC-53/1
	qb1R41071/R41071 Hk575-f Homo sapiens cDNA clone k575-f...	535	1.5e-72	4	hu-UNC-53/1
30	qb1N891041/N89104 K7846F Fetal heart, Lambda ZAP Expre...	451	7.3e-57	2	hu-UNC-53/1
	qb1R41073/R41073 Hk144-f Homo sapiens cDNA clone k144-f...	555	1.5e-36	1	hu-UNC-53/1
	qb1R15492/R15492 HH434-F Homo sapiens cDNA clone H434-F...	416	2.3e-29	2	hu-UNC-53/1
	qb1H09036/H09036 y196c11.r1 Homo sapiens cDNA clone 4...	438	9.4e-26	1	hu-UNC-53/2
	qb1W91567/W91567 MTA.C36.093.A MTA adult mouse thymus...	317	1.9e-17	2	ms-UNC-53/7
35	qb1W74400/W74400 zd62d10.r1 Soares fetal heart NbHH19...	243	2.2e-09	1	hu-UNC-53/1
	qb1AA003314/AA003314 mg56h10.r1 Soares mouse embryo NbME1...	141	0.54	1	

FIG. 7.

TBLASTN search of the Genbank sequence database with the 961 aminoacid ORF of cDNA 3b of hu-UNC-53/1. hu-UNC-53/1 forms a unique pair with Ce-UNC-53 (cosmid F45E10) compared to the rest of the database.

5

10

15

20

25

30

35

40

```

TBLASTN 1.4.3MP (26-March-1996) (Build 14:27:13 Apr 1 1996)
Query= tmpseq_1
      (961 letters)
Database: Non-redundant GenBank+EMBL+DDBJ+PDB sequences
          261,674 sequences; 371,416,172 total letters.

Sequences producing High-scoring Segment Pairs:

      Reading  High  Smallest
      Frame  Score  Sum
      P(N)    N

embI247810|CEF45E10  Caenorhabditis elegans cosmid F45E10 -2 158 2.3e-32 7
gbIM97501|HUMCLIP    Human cytoplasmic linker protein-... +3 83 0.47 1
embIX64838|HSRESTIN  H.sapiens mRNA for restin +1 83 0.47 1
gbIM58752|ECOMCRBC   E.coli_mcrB and mcrC genes, compl... +3 82 0.56 1
embI211582|SCNUF1G   S.cerevisiae nuf1 gene -1 82 0.61 2
embIX71297|SCSETRP4  S.cerevisiae spacer element -1 82 0.74 2
embIX54002|XLKINESIN X.laevis mRNA for kinesine +2 63 0.85 5
gbIU42409|DCU42409   Dictyostelium discoideum myosin h... +3 66 0.92 2
gbIU10399|YSC8082    Saccharomyces cerevisiae chromoso... +2 77 0.93 2
gbIU20810|ATU20810   Arabidopsis thaliana cytoskeleton... +1 77 0.95 2
gbIL07879|LEIKINLIKE Leishmania chagasi kinesin-like p... +2 78 0.95 1
gbIL03188|YSCINTANA  Saccharomyces cerevisiae integrin... +2 65 0.997 1
gbIU28372|YSCD9476   Saccharomyces cerevisiae chromoso... +3 82 0.9991 2
gbIM94362|HUMLAMBA   Human lamin B2 (LAMB2) mRNA, part... +1 75 0.9996 1
gbIM58337|VACHAGMA   Vaccinia virus hemagglutinin gene. -1 74 0.99995 1
  
```

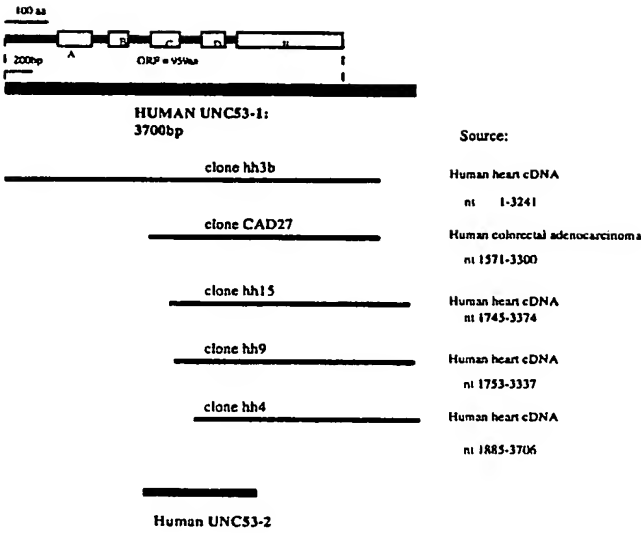


Figure 8

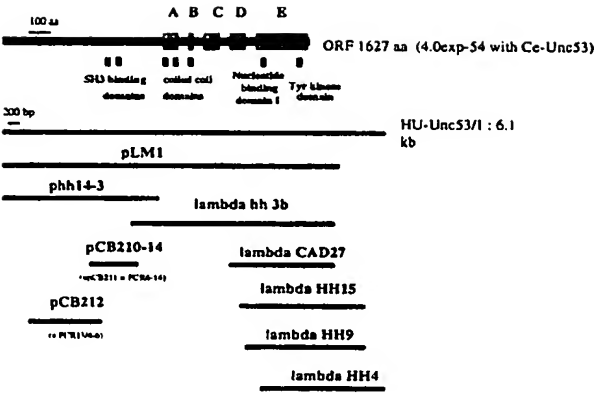


Figure 8a

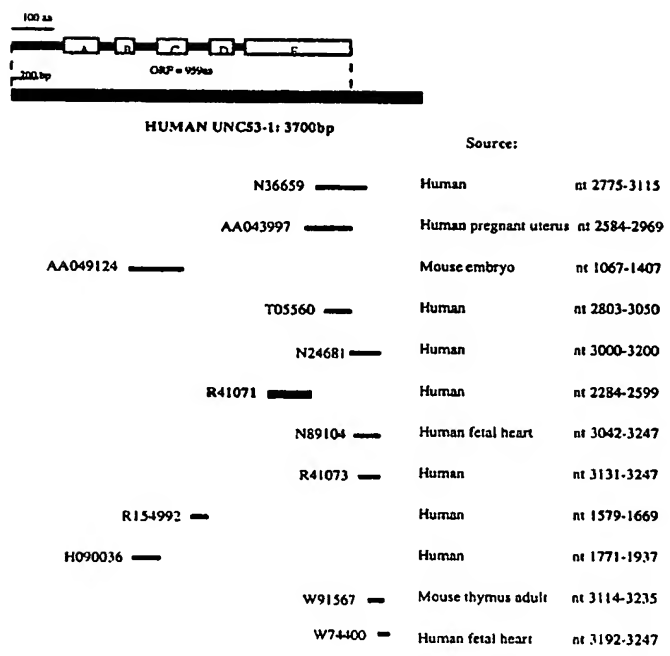


Figure 8b

FIG. 9a

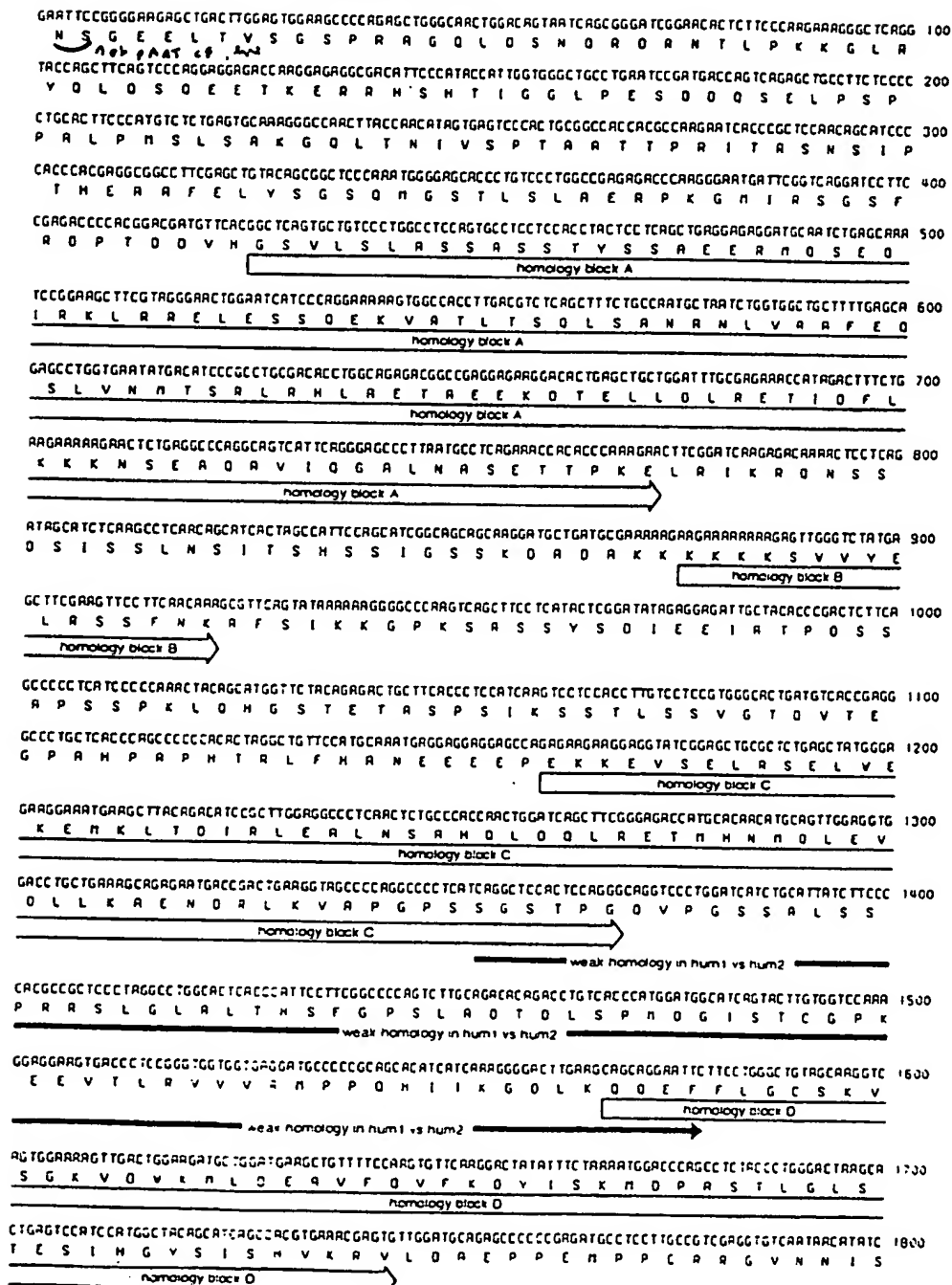


FIG. 9a CONTINUED.

AGTC TCCCTCAAGGCTCGAGGGGAATGGTGGTGACAGCCTGTATCCCAGGCCGATGTCGAGCACCTACATAAGCCCTCC TGCCT 1900
V S L K G L K E C C V D S L V F E T L I P K P M N O H V I S L L L

homology track E - pred nucleotide BD

AGGACACGGGGCCCC TCGTCTCTCTGGGCTCCAGGGGACGAGCACC TACTGACCACTCGCTTTGGCCGAGTACC TGGTGGAGCGCTCTGCGCCGTAGG 2000
K H R A L V L S S P S G T G K T Y L T M R L A E Y L V E R S G R E

homology track E - pred nucleotide BD

TCACAGAGGGCATCGTACGACCCTTACATGACCCAGCAGCTTGCAGGATCTGCACCTGTATCTTTCCARCCATTAGCCACGATAGACGGGAAAC 2100
V T E G I V S T F Y M H Q O S K O L Q L V L S N L A M O I D R E T

homology track E - pred nucleotide BD

AGGAATGGGGATGTGCCCC TGGTGATTCATGTGATGACC TGAGTGAGCGGCTCCATCAGTGAG TTGGTCATGGGGCCCC TCACC TGCAGTATCAT 2200
G I G O V P L V I L L D O L S E A G S I S E L V M G A L T C K Y H

homology track E - pred nucleotide BD

AATGTCCC TATATTAAGGTACCACTATCAGCTGTAAAATGACACCCCAACATGGC TTGCATCTGAGE TTCAGGATGTGACCTTC TCCACACAG 2300
K C P V I I G T T N O P V K A T P N H G L M L S F A A L T F S N H

homology track E - pred nucleotide BD

TGGAGCCAGCCAA TGGCTCC TGGTCTCTTACCTGAGGAGGAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAGAGAGCTGCTTGGGTGCT 2400
V E P A N G F L V A V L A R A K L V E S O S O I N A N K E E L L A V L

homology track E - pred nucleotide BD

CGAC TGGGTACCCAGGCTGTGGTATCATCTCCACACC TTCTTGAGAGGACAGCAGCCTCAGACTTCCCTCATCGGCCCTTGGTCTTCTGTGCTGTGCC 2500
O V V P K L V V M L H T F L E K H S T S O F L I G P C F F L S C P

homology track E - pred nucleotide BD

ATTGGCATTGAGGACTTCGGACCTGGTTCATGACC TGGGAACACA TCTATCA TCCCTATCTACGGAGGAGCCACAGGA TGGGA TAAGGTCCATG 2600
I G I E D F A T V F I O L V N N S I I P Y L O E G A K O G I X V H

homology track E - pred nucleotide BD

GACAGAARGCTGCTGGGGAGGACCAGTGGATGGGTCCGGGACACCTTCCC TGGCCATCAGCCCAACAGACCAATCAAGGCTGTACCACTTCCCC 2700
G O K A R A V E D P V E V V A D T L P V P S A O O O O S K L V H L P P

homology track E - pred nucleotide BD

ACCCACCGTGGGCCCTCACAGCA TTTCTCTCCTCCGAGGATAGGACAGTCAAGACAGCACCC CAAGTTC TCTGGAC TCAAGTCTCTGATGGCCATG 2800
P T V G P H S : A S P P E D A T V K O S T P S S L O S O P L A N A

homology track E - pred nucleotide BD

CTGCTGAACCTTCAGGAGCTGCCAACC TCAATGAGTCTCCAGATCGAGAACCA TCC TGGACCCCAACCTTCAGGACACACTTAAAGGGTTCGGCAATC 2900
L L K L Q E A A N V I E S P O R E T I L O P N L Q A T L

homology track E - pred nucleotide BD

ACTGTERECCCGGACAGCAGACGCTGGATCAGCTATCTTAGCTCC TCC TC TCCC TC TCC TC TTTCAGAGCAC TGGCTTCACGCCCCAGGAGGAGA 3000

3' untranslated trailer

ACAGGAGGGAGGAGGAGTGAAGAGGGGGGACAGGTTCTTGGTGC GTGACTTTGAGACTTCC TAGGAGGAGTGGTGGGG TGGCGTTTGGGAC TTG 3100

3' untranslated trailer

TGCCCC TAACACATTTAC TGGCTCTCTCAATGAC TTGGGGAAAGATGATTCTGGGCTTTCCCTTGACTCTTTGTTCAATTACAACTCTCTGG 3200

3' untranslated trailer

CTTTCTGGGGAGGGGTTCAGAAACCTTAAACATCGACAGTTCCTAATGATCTTCACAGCAGCCCTGAGAGAGACAGTCTTTGTAGGGAGATCTG 3300

3' untranslated trailer

GGGAGGAGGAGGAGCTCTCAGATTTCTTCAAGACCTTCEATTCATCACCCTGCCACAACTCC TCCCCAGAGATCTGGCTGGAGCCCCAGAA 3400

3' untranslated trailer

AGAGACATGTGGTTTAAAAATCTTAAATCACTGTAAAGGTAAAAATGAAAAACAAAAACAGCAACCAACAAAAACAATGGAAAGATGAA 3500

3' untranslated trailer

GC TGGAGAGGAGGAGGAGCTCTCAGATTTCTTCAAGACCTTCEATTCATCACCCTGCCACAACTCC TCCCCAGAGATCTGGCTGGAGCCCCAGAA 3600

3' untranslated trailer

TCACCAACCAAAAAATACCTTCAAGCTTTCAGGAGAGCTTACAGCTCTGCTTCTTACCC TCTAATTTAACTATGACCGGAATTCAGCTTTGGAC 3700

3' untranslated trailer

ITAC 3705

Fig. 9b

Tuesday, 18 November 1997 10:33

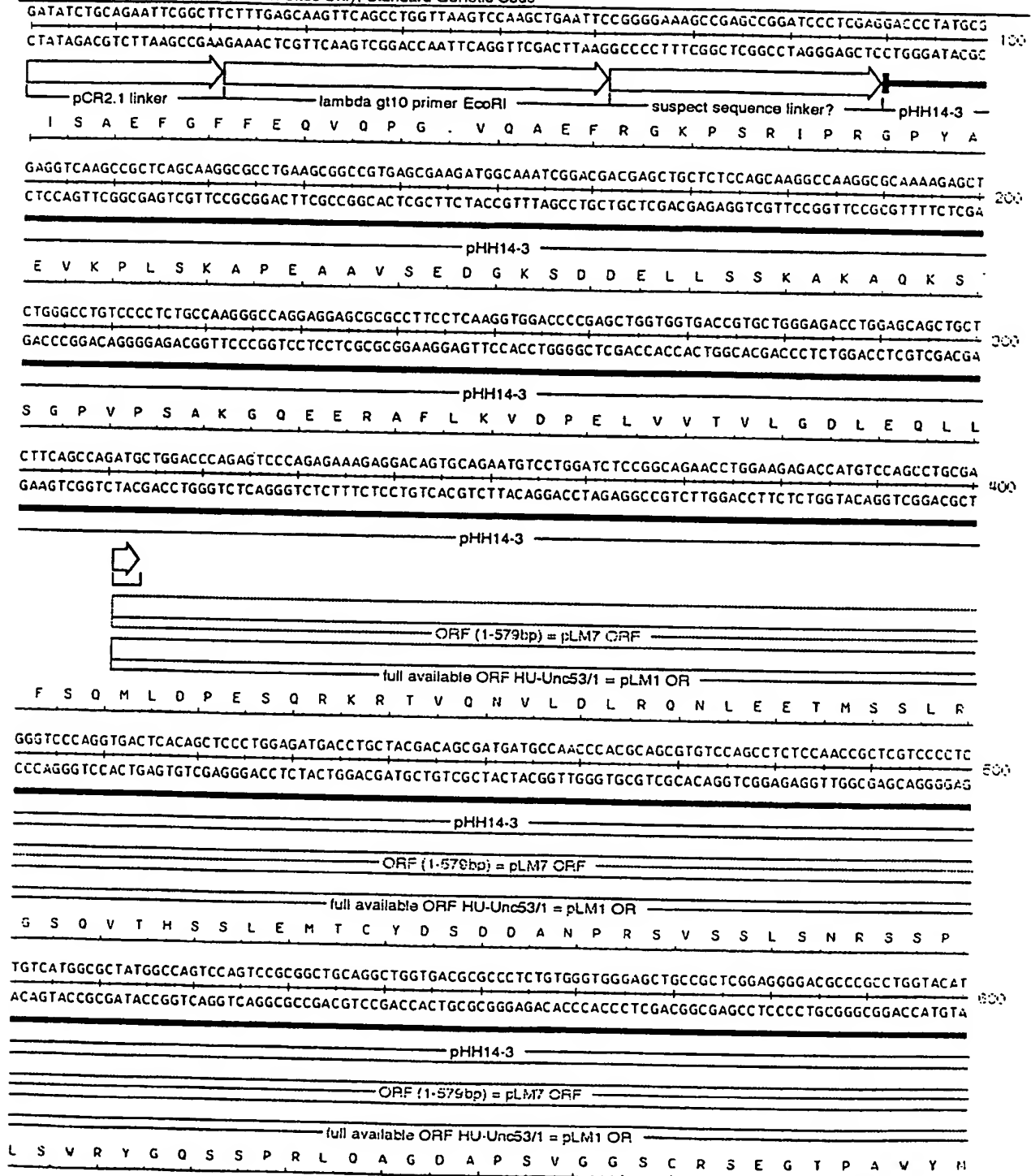
lu-Unc53/1 seq (1 > 8013) Site and Sequence

Enzymes: 60 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Fig 9 14 pag



Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 7

GCACGGCGAACGGGCCAC TACTCCCACACCATGCCATGCGCAGCCCCAGCAAGCTCAGCCATATCTCCCGCTGGAGCTGGTCTGAATCCCTGSACTCG 700
CGTGCCGCTTGCCCGGGTGATGAGGGGTGGTACGGGTACGGCTCGGGGTCGTTTCAGTTCGGTATAGAGGGCGGACCTCGACCAGCTTAGGGACCTGAGC

————— pHH14-3 —————
————— ORF (1-579bp) = pLM7 ORF —————
————— full available ORF HU-Unc53/1 = pLM1 OR —————
H G E R A H Y S H T M P M R S P S K L S H I S R L E L V E S L D S

GATGAGGTGGACCTCAAGTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACATCACTACCGGCTGGGATGAAAGCA 800
CTACTCCACCTGGAGTTTCAAGCCGATGTACTCGCTGTCAGTGGAGTACCCGTTCTGGTACTGCCCTCTACTACTGTAGTGATGGCCGACCTTACTTTCGT

————— pHH14-3 —————
————— ORF (1-579bp) = pLM7 ORF —————
————— full available ORF HU-Unc53/1 = pLM1 OR —————
D E V D L K S G Y M S D S D L M G K T M T E D D D I T T G V D E S

GCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAATCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCCACTGC 900
CGAGGTAGTCATCACCTGAGTCGCTACGGAGTCTGTTAGAGTCAAGTCTTCTTAAGTTACGGTCGAGGAGTGAGTTGAGGGAGGGTTCATGAGGGTGAGC

————— pHH14-3 —————
————— pCB212 —————
————— ORF (1-579bp) = pLM7 ORF —————
————— full available ORF HU-Unc53/1 = pLM1 OR —————
S S I S S G L S D A S D N L S S E E F N A S S S L N S L P S T P T A

TTCTCGCAGGAACCTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGCAGAAAGTGGGCTGAGCTGGTTAGTGAATCAGAGGAGAAAGCC 1000
AAGAGCGTCCTTGAGTTGTATCACGATGCGGTGCTGAGTCTCTTCGCGAGTGACCGTCTTTCACCCGACTCGACCAATCACTTAGTCTCTCTTTGGG

————— pHH14-3 —————
————— pCB212 —————
————— full available ORF HU-Unc53/1 = pLM1 OR —————
S R R N S T I V L R T D S E K R S L A E S G L S W F S E S E E I A

CCTAAAAAAGTGGAGTACGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCTGAGAGCTGTGATGATTCATCCAAAGS 1100
GGATTTTTTGACCTCATGCTGTCACCATCGGACTTCTACCTTGGACCTGAAGATTACCGGCTCCCTCGCCGGACTCTCGACACTACTAAGTAGGTTCC

————— pHH14-3 —————
————— pCB212 —————
————— full available ORF HU-Unc53/1 = pLM1 OR —————
P K K L E Y D S G S L K M E P G T S K V R R E R P E S C D D S S I

Tuesday, 18 November 1997 10:33

Fig 9b

Page 1

fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

GTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCCCTGGTTCCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCCATCACTCACACAGC
CACCTCTTGACTTTTTCGGGTAGTCGGACCCGGTGGGACCAAGGGACTTCTCCCGTCTGGGGTGGACACCGACATTGAAGGGGGTAGTGAGTGTGTCTG 120

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

G G E L K K P I S L G H P G S L K K G K T P P V A V T S P I T H T A

CCAGAGTGGCCTCAAAGTCGCAGGCAAACTGAGGGCAAGCTACAGACAAGGGTAAGCTTGCAAGTGAAGAATACTGGGCTCCAACGCTCCCTCTCTGA7
GGTCTCACGGGAGTTTCAGCGTCCGTTTGGACTCCCGTTTCGATGCTGTTCCTTCGAACGTCACCTTCTATGACCCGAGGTTCGAGGAGGAGACTA 130

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

O S A L K V A G K P E G K A T D K G K L A V K N T G L O R S S S D

GCTGGTCGGGACCGCCTGAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTTGGCTACAAGAAGCCTCCCTCTGCCA
CGACCAGCCCTGGCGGACTCACTACGATTCTTCGGGGGAGCCCGTAACGAGCGGGGAGGTGAAGCCCTAGGAAACCGATGTTCTTCGGAGGAGGACGGT 140

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

A G R D R L S D A K K P P S G I A R P S T S G S F G Y K K P P P A

CAGGCACAGCCACTGTATGCAAACTGGTGGTTCAGCCACTCTCAGCAAGATCCAGAAGTCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAAGAC
GTCCGTGTGGGTGACAGTACGTTTGACCACCAAGTCGGTGAGAGTCTGTAGGTCTTCAGGAGTCCGTAGGGACAGTTCGGTCATTACCCGCGTCTCT 150

pHH14-3

pCB212

full available ORF HU-Unc53/1 = pLM1 OR

T G T A T V M O T G G S A T L S K I O K S S G I P V K P V N G R P T

TAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCTTGCTCTGGAGCCCGTTCTAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAAGTCAAGT
ATCGAATCTACAAAGGTTGTACAGTCTCGGTCCTAAGGACCGAGGACCTCGGGCAAGATTGTAGGTTCATGGCGTCGGACGGGGCCGGTTCAGTTCA 160

pHH14-3

pCB212

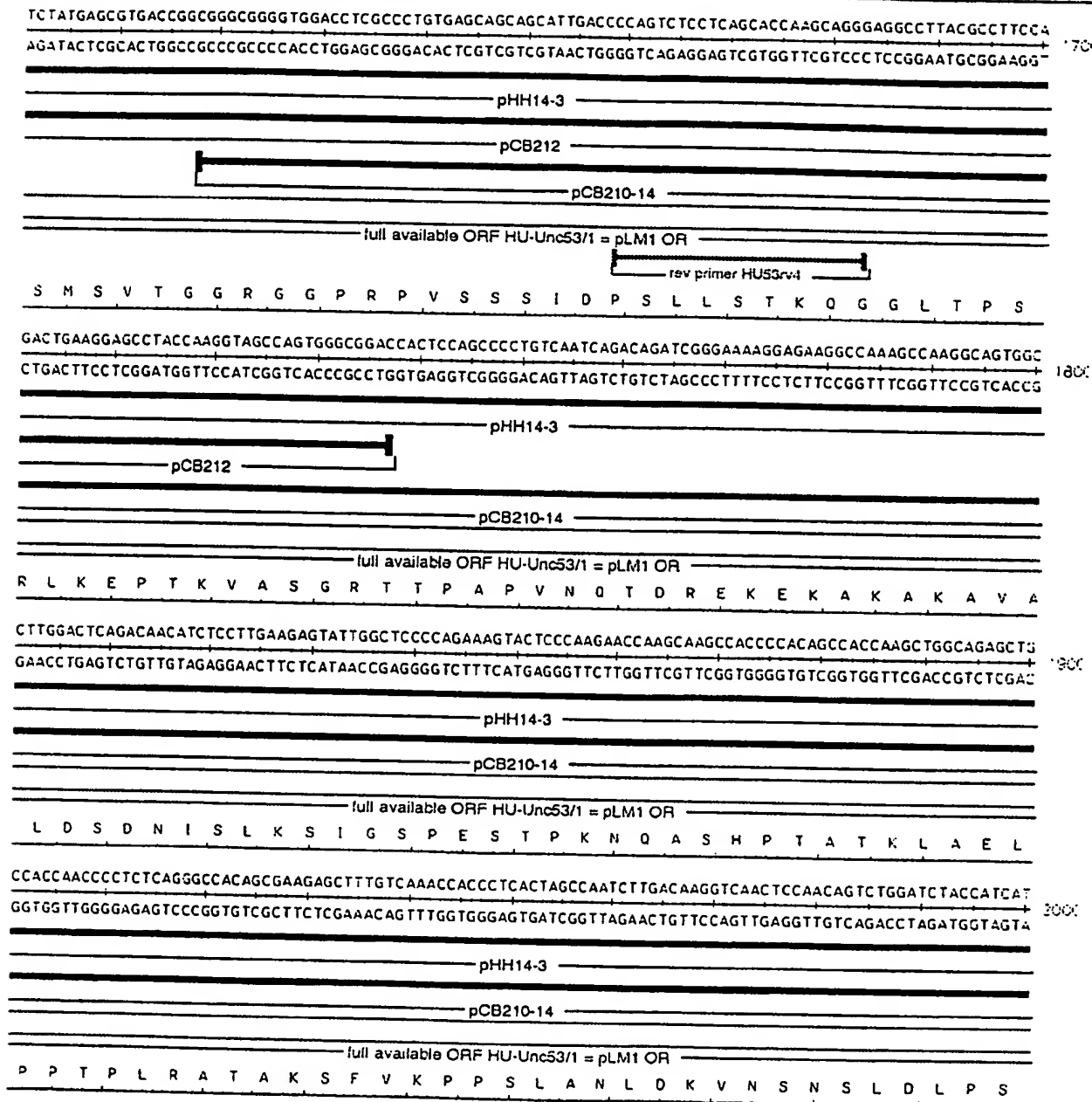
full available ORF HU-Unc53/1 = pLM1 OR

S L D V S N S A E P G F L A P G A R S N I Q Y R S L P R P A K S S

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 4



Tuesday, 18 November 1997 10:33

fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 5

CCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCTGCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTACCCCCAGTCCGGCACC
GGTCACTATGGTGGGTACGAAGTTTCCAGGGTCTAGACGTACGATGTTGAGTCGTAGACCCCGGAGAGGGAAGGACGAAGTGGGGGTCAGGCCGTGG 210

————— pHH14-3 —————
————— pCB210-14 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————
S S D T T H A S K V P D L H A T S S A S G G P L P S C F T P S P A P

CATCCTCAATATTAACTCAGCCAGCTTCTCCAGGGCCTGGAGCTAATGAGTGGTTTCAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACCTCAGGC
GTAGGAGTTATAATTGAGTCGGTCTGAAGAGGGTCCCGGACCTCGATTACTCACCAAAGTCACACGGTTTTCTCTGGGCGTACATGGGGTTTGAGAGTCCG 220

————— pHH14-3 —————
————— pCB210-14 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————
I L N I N S A S F S Q G L E L M S G F S V P K E T R M Y P K L S G

CTGCACAGGAGCATGGAGTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCACCCACCTGCTCCCCCTGTGTCTC
GACGTGTCTCGTACC TCAGGGAGGTCTACGGTTACTCGGAGGGGTACGGAAGGGGTCGTCATGAGGGCAGGGGTGGGGTGGACGAGGGGGACGACGAS 230

————— pHH14-3 —————
————— pCB210-14 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————
L H R S M E S L Q M P M S L P S A F P S S T P V P T P P A P P A A

CCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGGGAAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAAAGG
GGTGTCTTCTCTCTGCCTTCTCGACTGAACCTACCTTCGGGGTCTCGACCCGTTGACCTGTCAATTAGTCGCCCTAGCCTTGTGAGAAGGGTTCTTTCC 240

————— pHH14-3 —————
————— pHH3b —————
————— pCB210-14 —————

————— full available ORF HU-Unc53/1 = pLM1 OR —————
————— rev primer HU53rv3 —————
————— rev primer HU53rv2 —————
————— peptide B72828H —————

P T E E E T E E L T V S G S P R A G Q L D S N Q R D R N T L P K K G

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 4

GCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATACCATGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCT 250
CGAGTCCATGGTCAAGTCAGGGTCTCTCTGGTTCCTCTCCGCTGTAAGGGTATGGTAACCAACCCGACGGACTTAGGCTACTGGTCAGTCTCGACGGAA

pHH14-3

pHH3b

rev primer HU53rv1

full available ORF HU-Unc53/1 = pLM1 OR

L R Y Q L Q S O E E T K E R R H S H T I G G L P E S D D Q S E L P

TCTCCCCCTGCACCTCCCATGTCTCTGAGTGCAAGGGCCAACTTACCAACATAGTGAGTCCCAGTCCGCGCCACCACGCCAAGAATCACCCGCTCCAACA 260
AGAGGGGGACGTGAAGGGTACAGAGACTCACGTTTCCCGGTGAATGGTTGTATCACTCAGGGTGACGCCGGTGGTGCAGTTCTTAGTGGGCGAGGTTGT

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S P P A L P M S L S A K G Q L T N I V S P T A A T T P R I T R S N

GCATCCCCACCCACGAGGCGGCC TTCGAGCTGTACAGCGGCTCCCAATGGGGAGCACCTGTCCCTGGCCGAGAGACCCAAGGAATGATTGGGTCAGG 270
CGTAGGGGTGGGTGCTCCGCCGAAGCTCGACATGTGCCGAGGGTTTACCCTCGTGGGACAGGGACCGGCTCTCTGGGTTCCCTTACTAAGCCAGTCC

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S I P T H E A A F E L Y S G S Q M G S T L S L A E R P K G M I R S G

ATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGCTGCTCCCTGGCCTCCAGTGCTCTCCACCTACTCTCAGCTGAGGAGAGGATGCAATCT 280
TAGGAAGGCTCTGGGGTGCTGCTACAAGTGCCGAGTCACGACAGGGACCGGAGGTACGGAGGAGGTGGATGAGGAGTCGACTCTCTCTACGTTAGA

pHH14-3

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S F R D P T D D V H G S V L S L A S S A S S T Y S S A E E R M Q S

GAGCAATCCGGAAGCTTCGTAGGGAAC TGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGCCAATGCTAATCTGGTGGCTGCTT 290
CTCGTTAGGCCCTTCAAGCATCCCTTGACCTTAGTAGGGTCTTTTACCAGGTGGAACGCAGAGTCGAAAGACGGTTACGATTAGACCACCGACGAA

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

E O I R K L R R E L E S S O E K V A T L T S O L S A N A N L V A A

Tuesday, 18 November 1997 10:33
fig Hu-Unc53/1 seq (1>6013) Site and Sequence

Fig 95

Page 7

TTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCGACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTGCGAGAAACCATAGA
AACTCGTCTCGGACCACCTTACTGTAGGGCGGACGCTGTGGACCGTCTCTGCCGGCTCCTCTTCCGTGACTCGACGACCTAAACGCTCTTTGGTATCT 300

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F E Q S L V N N T S R L R H L A E T A E E K D T E L L D L R E T I D

CTTCTGAAGAAAAAGAACTCTGAGGCCAGGCGAGTCATTAGGAGGCCCTTAATGCCTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAAC
GAAAGACTTCTTTTCTTGAGACTCCGGGTCCGTGAGTAAGTCCCTCGGGAATTACGGAGTCTTTGGTGTGGGTTTCTTGAAGCCTAGTCTCTGTTTGT 310

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

F L K K K N S E A Q A V I Q G A L N A S E T T P K E L R I K R Q N

TCCTCAGATAGCATCTCAAGCCTCAACAGCATCAC TAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAAAAAGAGAAAAAGAGTTGGG
AGGAGTCTATCGTAGAGTTCGGAGTTGTCGTAGTGATCGGTAAGGTCGTAGCCGTCGTCGTTCCACGACTACGCTTTTCTCTTTTTTCTCAACCC 320

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S D S I S S L N S I T S H S S I G S S K D A D A K K K K K K S V

TCTATGAGCTTCGAAGTTCTTCAACAAAGCGTTCAGTATAAAAAAGGGGCCAAGTCAGCTTCTCATACTCGGATATAGAGGAGATTGCTACACCCGA
AGATACTCGAAGCTTCAAGGAAGTTGTTTCGCAAGTCATATTTTTCCCGGGTTCAAGTCAAGGAGTATGAGCCTATATCTCTCTAACGATGTGGCT 330

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

V Y E L R S S F N K A F S I K K G P K S A S S Y S D I E E I A T P D

CTCTTCAGCCCCCTCATCCCCAAACTACAGCATGGTTCACAGAGACTGCTTCACCCCTCCATCAAGTCTCCACCTTGCTCTCCGTGGGACTGATGT
GAGAAGTCGGGGGAGTAGGGGTTTGATGTCTACCAAGATGTCTCTGACGAAGTGGGAGGTAGTTCAGGAGGTGGAACAGGAGGCACCCGTGACTACG 340

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

S S A P S S P K L Q H G S T E T A S P S I K S S T L S S V S T D V

Tuesday, 18 November 1997 10:33

file: hu-Unc53/1 seq (1 > 6013) Site and Sequence

Page 8

Fig 9b

ACCG .GCCCTGCTCACCCAGCCCCACACTAGGCTGTTCATGCAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGAGCTGCGCTCTGAGG
TGGCTCCCGGGACGAGTGGGTCGGGGGGTGTGATCCGACAAGGTACGTTACTCTCTCTCGGTCTCTTCTCTCTCCATAGCCTCGACGCGAGACTCG 3500

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

T E G P A H P A P H T R L F H A N E E E E P E K K E V S E L R S E

TATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGGAGGCCCTCAACTCTGCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTT
ATACCTCTCTCTTTACTTCGAATGTCTGTAGGCGAACCTCCGGGAGTGTAGACGGGTGGTTGACCTAGTCGAAGCCCTCTGGTACGTGTGTACGTCAA 3600

U2 ORF = pCB251 ORF

pHH3b

peptide B72527H

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

L V E K E M K L T D I R L E A L N S A H Q L D Q L R E T M H N M Q L

GGAGGTGGACCTGCTGAAAGCAGAGAATGACCGAC TGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCCTGGATCATCTGCATTA
CCTCCACCTGGACGACTTTCGCTCTTACTGGCTGACTTCATCGGGTCCGGGGAGTAGTCCGAGGTGAGGTCCCGTCCAGGGACCTAGTAGACGTAAT 3700

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

E V D L L K A E N D R L K V A P G P S S G S T P G Q V P G S S A L

TCTTCCCCACGCCGCTCCCTAGGCC TGGCACTCACCCATTCCTTCGGCCCCAGTCTTGACAGACAGACCTGTACCCATGGATGGCATCAGTACTTGTG
AGAAGGGGTGCGGCGAGGGATCCGGACCGTGAGTGGTAAGGAAGCCGGGGTCAGAACGTCTGTGCTGGACAGTGGGTACCTACCGTAGTCATGAACAC 3800

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

S S P R R S L G L A L T H S F G P S L A D T O L S P M D G I S T C

Tuesday, 18 November 1997 10:34

file 'u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 9

GTCCAAAGGAGGAAGTGACCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCTGGGCTTAG
CAGGTTTCTCTCTTCACTGGGAGGCCACCACCACCTCTACGGGGCGCTCGTGTAGTAGTTTCCCTGAAC TTCGTCGTCCTTAAGAAGGACCCGACATC

380

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

G P K E E V T L R V V V R M P P O H I I K G D L K Q Q E F F L G C S

CAAGGTCAGTGGAAAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAGGACTATATTTCTAAAATGGACCCAGCCTCTACCCTGGGA
GTTCCAGTCACCTTTTCAACTGACCTTCTACGACCTACTTCGACAAAAGGTTACAAGTTCC TGATATAAAGATTTTACCTGGGTGGGAGATGGGACCTT

400

U2 ORF = pCB251 ORF

pHH3b

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

K V S G K V D V K M L D E A V F Q V F K D Y I S K M D P A S T L G

CTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATA
GATTCGTGACTCAGGTAGGTACCGATGTCGTAGTCGGTGACATTTGCTCACAACCTACGTCTCGGGGGGCTCTACGGAGGAACGGCAGCTCCACAGTTAT

420

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L S T E S I H G Y S I S H V K R V L D A E P P E M P P C R R G V N

ACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAAATGCGTCGACAGCCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCT
TGATATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGCAGCTGTCGGACCAACAAGCTCTGCGACTAGGGGTTCGGCTACTACGTCGTGATGTATTGGGA

440

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

H I S V S L V G L K E K C V D S L V F E T L I P K P M M Q H Y I S L

Tuesday, 18 November 1997 10:34

fig. 14-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 10

CCCTCTGAGCACCAGCGCCTCGTCCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGCTGGAGCGCTCTGGC
GGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCCGGGTCGCCGTGCCGTCTTGGATGGACTGGTTAGCGAACCGGCTCATGGACCACTCGCGAGACCG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

L L K H R R L V L S G P S G T G K T Y L T N R L A E Y L V E R S G

CGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCCAGATAGACC
GCACTCCAGTGCTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGTCGTCAGAACGTTCTTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E V T E G I V S T F N M H Q Q S C K D L Q L Y L S N L A N Q I D

GGGAAACAGGAATTGGGGATGTGCCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGC TCCATCAGTGAGTTGGTCAATGGGGCCCTCACTGCA
CCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTACTGGACTCACTTCGTCCGAGGTAGTCACTCAACCAGTTACCCCGGGAGTGGACGT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R E T G I G D V P L V I L L D D L S E A G S I S E L V N G A L T C

Tuesday, 18 November 1997 10:34
fig. 4u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 17

GTATCA:AAATGTCCCTATATTATAGGTACCACCAATCAGCC TGTAAAAATGACACCAACCATGGCTTGCAC TTGAGCTTCAGGATGTTGACCTTCCTC
CATAGTATTACAGGGATATAATATCCATGGTGGTTAGTCGGACATTTTACTGTGGGTGGTACCGAAGCTGAAC TCGAAGTCCTACAAC TGGAAAGAGG 4601

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

peptide B72626H

Y H K C P Y I I G T T N Q P V K M T P N H G L H L S F R M L T F S

AACAACGTGGAGCCAGCCAATGGCTTCCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGC TGCTTC
TTGTTGCACCTCGGTTCGGTTACCGAAGGACCAAGCAATGGACTCCTCCTTCGACCATCTCAGTCTGTCGCTGTAGTTACGGTTGTTCTTCGACGAAS 4701

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

N N V E P A N G F L V R Y L R R K L V E S D S D I N A N K E E L L

GGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTG
CCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGAAGGAACCTTCGTGTCGTGGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAS 4801

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

R V L D V V P K L V Y H L H T F L E K H S T S D F L : G P C F F L S

Tuesday, 18 November 1997 10:34
fig. HU-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 12

GTGTCCCATTTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACCTCTATCATTCCTTATCTACAGGAAGGAGCCAAAGGATGGGATAAAG
CAGAGGGTAACCGTAACCTCTGAAGGCCCTGGACCAAGTAAGTGGACACCTTGTGAGATAGTAAGGGATAGATGTCTTCTCGGTTCCCTACCTTATTTCT

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

C P I G I E D F R T V F I D L V N N S I I P Y L Q E G A K D G I I

GTCCATGGACAGAAAGCTGTCTGGGAGGACCCAGTGGAAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACT
CAGGTACCTGTCTTTCGACGAACCTCTCTGGGTCACCTTACCCAGGCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGG

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

V H S O K A A V E D P V E W V R D T L P V P S A O O D O S K L Y H

TGCCCCCACCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGA
ACGGGGGTGGGTGGCACCCGGGAGTGTGTAACGGAGTGGAGGGCTCTATCTGTGAGTTTCTGTGCGTGGGGTTCAAGAGACCTGAGTCTAGGAGACTA

U2 ORF = pCB251 ORF

pHH3b

U4 ORF = pCB201 ORF

full available ORF HU-Unc53/1 = pLM1 OR

U3 ORF = pLM5 ORF

pHH15

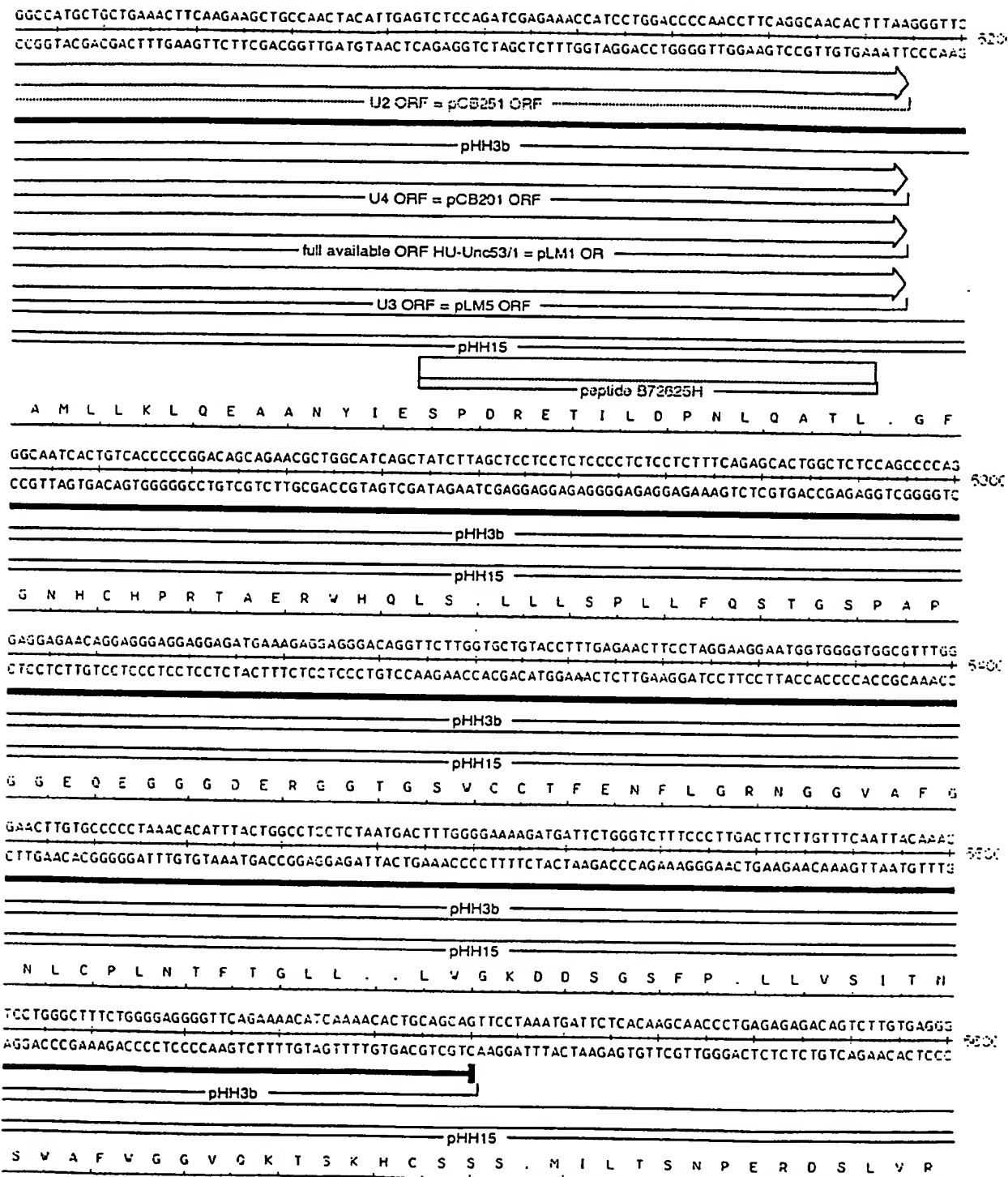
L P P P T V G P H S I A S P P E D R T V K D S T P S S L D S O P L I

Tuesday, 18 November 1997 10:34

fig. 4u-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 13



Tuesday, 18 November 1997 10:34
fig 1Hu-Unc53/1 seq (1 > 6013) Site and Sequence

Fig 9b

Page 4

AGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCACAGACCCTTCCCAATCCATCACCCTGCCAACAACTCTCCCCCAGAGATCTGGCTGGAGC
TCTAGACCCCTCCGCTCTTCGAGGAGTCTAAAAGAGTGCTGGGAAGGGTTAAGGTAGTGGTGACGGTTGTTGAGGAGGGGCTCTAGACCGACCTCG 570

—————pHH15—————→

E I V G R Q E A P Q I F S Q T L P N S I T T A N N S S P R D L A G A
CCAGAAAAAGAAGCATGTGGTTTAAAAATGTTTAAATCAATCTGTAAAGGTAAAAATGAAAAACAAAAACAAGCAACAAACAAAAACAATGGAAA 580
GGTCTTTTCTTCGTACACCAATTTTTACAAATTTAGTTAGACATTTCCATTTTACTTTTTGTTTTGTTCTGTTTGTGTTTTGTTTACCTTT
O K K K H V V . K M F K S I C K R . K . K N K N K Q T N K K Q W I
AGATGAAGCTGGAGAGAGAGGAACCAAGTTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAAC 590
TCTACTTCGACCTCTCTCCTTGGTCAACGGTTCCATCTCTCGACGGCGAGGACGGGAGACCTACTGTATCCCTGTAGTTGTCTGCCGACGGTTGG
R . S V R E R N Q L P R . R A A R S C P L D D I G D I N K T A A N
TGAGAAGTCACCAAAACCAAAAATAACCTTACAGCCTTCAGGGAAGACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCACCGGAATTCAG 600
ACTCTTCAGTGGTTTGGTGTGTTTTATTGGAATGTCGGAAGTCCCTTCTGATGGTCGAGACAGAAAGATGGGAGATTAAATTGTTACGTGGCCTTAAGTC
linker? —
L R S H Q T T K I T L Q P S G K D Y Q L C L S T L . F N N A P E F S
CTTGGACTTAACC
GAACCTGAATTGG 6013
—————→
linker? —
L D L T

FIG. 10.

GGCACGAGGCATCCTCTGTGGGCACCGAGGTCACCGAGACCCCTGCTCATTCAGTCCCCCACACTAGACT 70
linker ? open reading frame

R H E A S S V G T E V T E T P A H S V P H T R L
GTTCCAAGCCAATGAAGAGGAGGAGCCAGAGAAGAAGGAGGTATCAGAACTGCGCTCTGAACTATGGGAA 140
open reading frame

F Q A N E E E E P E K K E V S E L R S E L V E
AAAGAGATGAAGCTCACGGATATCCGGTTGGAGGCCCTCAACTCTGCCACACAGCTGGACCAGCTTCGGG 210
open reading frame

K E M K L T D I R L E A L N S A H O L D O L R
AGACCATGCACAATATGCAGTTGGAGGTGGACCTGCTGAAAGCAGAGAATGACCGGCTGAAGGTGCCCC 280
open reading frame

E T M H N M Q L E V D L L K A E N D R L K V A P
CGGCCCCCTCTCAGGCTGCACTCCAGGGCAGGTCCTGGGTCATCGGCTCTGTCGTCCTCCCTCGACGTTCC 350
open reading frame

G P S S G C T P G O V P G S S A L S S P R F S
CTGGGCCCTTGCACTCAGCCATCCTTTTCAGTCTTAGTCTCACAGACACAGACCTCTCACCCATGGATGGCA 420
open reading frame

L G L A L S H P F S P S L T D T G L S P M C G
TCAGCACCTGTGGTTCAAAGGAAGAGGTGACCCTGCGGGTGGTGGTCCGGATGCCGCCCCAGCACATCAT 490
open reading frame

I S T C G S K E E V T L R V V V R M P P D H I
CAAAGGGGACTTAAAGCAGCAGGAGTTCTTCTGGGTTGCAGCAAGGTGAGTGGCAAGTTGACTGGAG 560
open reading frame

K G D L K Q Q E F F L G C S K V S G K V D K
ATGCTGGATGAAGCCGTTTTCCAAGTGTTCAGGACTACATTTCTAAATGGAGCCAGCCTCAAGCTGG 630
open reading frame

M L D E A V F Q V F D Y I S K M D F A S T L
GACTGAGCACTGAGTCCATACATGGCTATAGCCTCAGCCACGTGAACGAGTGCTGGATGCTGAGGCCCTTC 700
open reading frame

G L S T E S I H G Y S L S H V K R V L D A E F P

FIG. 10 CONTINUED.

AGAGATGCCCTCCTTGCCGCCGAGGTGTCAATAACATATCAGTCGCTCTCAAAGGTCTGAAAGAGAAGTGT 770
open reading frame

E M P P C R R G V N N I S V A L K G L K E K C
GTCGACAGCCTGGTGTTCGAGACGCTTATCCCAAGCCCATGATGCAGCACTACATCAGCCTCCTGCTCA 840
open reading frame

V D S L V F E T L I P K P M M Q H Y I S L L L
AGCACCGGCGCCTGGTGCTCTCCGGCCCCAGTGGCACCGGCAAGACCTACTTGACCAATCGGCTAGCCGA 910
open reading frame

K H R R L V L S G P S G T G K T Y L T N P L A E
GTACCTGGTGGAGCGCTCCGGCCGCGAGGTACGGATGGCATCGTCAGCACTTTCAACATGCACCAGCAG 980
open reading frame

Y L V E R S G R E V T D G I V S T F N M H Q Q
TCTTGCAAGGATCTGCAACTGTACCTCTCCAACCTAGCCAACCAGATAGACCGGGAAACAGGGATAGGGG 1050
open reading frame

S C K D L Q L Y L S N L A N Q I D R E T G I G
ATGTGCCCTTGGTGATCCTCCTGGATGATCTGAGTGAAGCAGGCTCCATCAGTGAGCTGGTCAATGGGGC 1120
open reading frame

D V P L V I L L D D L S E A G S I S E L V N G A
CCTCACCTGCAAGTATCACAAATGTCCCTACATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCC 1190
open reading frame

L T C K Y H K C P Y I I G T T N Q P / K M T P
AACCATGGCTTGCACITGAGCTTCAGGATGCTGACCTTCTCGAACAATGTGGAACCAGCCCATGGCTTTG 1260
open reading frame

N H G L H L S F R M L T F S N N V E P A N G F
TGGTCCGTTACCTGCGGAGGAAGTTGGTAGAGTCAGACAGTGACGTCAATGCTAACAGGGAAGAGCTTCT 1330
open reading frame

L V R Y L R R K L V E S D S D V N A N I E E L L
TCGGGTGCTGGACTGGGTGCCCAAGCTGTGGTATCACCTCCACACCTTCTGGAGAAAGCAAGCACCTCG 1400
open reading frame

R V L D V V P L V V H L H T F L E . H S : S

FIG. 10 CONTINUED.

AACAAAAACAAACAAACCAACTACAGCAGTTCCAAGCTCGTTCTCACAAACACCTCTGAGACAGTCACAT 2240
3' untranslated
T K T N K P T T A V P S S F S Q T P L R Q S H
GTGGGCAAACTAAGGGAGGCAGGAAGCTCTACAGACTTTCTTGCAAACCTTCCCAGTTCTGTGCGACAC 2310
3' untranslated
V G K S K G G R K L Y R L S C K P F P V L S T
TGCCAACAACCTCCCCGCCAGAGACCTGGCCAGAGCCAAGAAAAGAGAAGCATGTGGTTTAAACAGAAAAA 2380
3' untranslated
L P T T S P P E T V P E P R K E K H V V . Q K N
CAAAACAAAACAAAACAAAATATATGTGTAAATCAACCTGTAGAAGGTAAAAACGGCAATGGAAAAGA 2450
3' untranslated
K T K Q N K K Y M C K S T C R R . K R Q V K R
TGAAGCTGGAAGGAGGGGCCAGTTGCCAAGATGGAACGAGAGCTGCCAGATCTTGCCCTCTGGATGACA 2520
3' untranslated
. S V K E G P S C Q D G T R A A R S C L L D D
AGAGGGGACATTGCAAGATGGCTGCCAGTCTAAAACGTCACCAGACCACAAGAGTAACATCACAGCCTTC 2590
3' untranslated
K R G H C K M A A S L K R H Q T T R V T S Q P S
GAAGAAAGGCCACAAGCTGTCTTCTTGCCCTCTAACTGAACATGCATGAAAAGTCAATAAACCCCTACTTT 2660
3' untranslated
K K G H K L S F C P L T E H A . K V N K P Y F
TTAATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAATTCGCGGCCGC 2709
polyA tail + linker
L I F K K K K K K K K K F P R P

FIG. 11a.

AAGCTTGGCAGGAGGCTCGTGCCAAGCTGAGACCGTCATGCAGCTCCGAAATGAGTTAAGAGACAAGGA 70
LINKER ? open reading frame
A V H E A S C Q A E T V M Q L R N E L R D K E
GATGAAGCTGACAGATATCCGCTTAGAAGCTCTCAGTTCTGCCACCAGCTGGACCAGCTCCGGGAGGCC 140
open reading frame
M K L T D I R L E A L S S A H O L D O L R E A
ATGAACAGGATGCAGAGTGAAATAGAGAAGCTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAG 210
open reading frame
M N R M Q S E I E K L K A E N D R L K S E S Q
GCAGTGGCTGCAGCCGGGCTCCTTCCCAAGTGTCATCTCTGCCTCCCGAGGCAGTCCATGGGCCCTCTC 280
open reading frame
G S G C S R A P S Q V S I S A S P R Q S M G L S
CCAGCACAGCTTGAACCTCACTGAGTCAACCAGCCTGGACATGTTGCTGGATGACACTGGTGAATGCTCG 350
open reading frame
Q H S L N L T E S T S L D M L L D D T G E C S
GCTCGGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTGAGCTTTTCAGGAGGAAATGAAGTGAAGGAGG 420
open reading frame
A R K E S G R H V K I V V S F Q E E M K V K E
ATTCAGACCACACCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGAAGACGAAGTGGGATGTGCTCGA 490
open reading frame
D S R P H L F L I G C I G V S G K T K V D V L D
TGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCATGTCGACCCAGTGAGTCAGCTAGGGCTGAAT 560
open reading frame
G V V R R L F K E Y I I H V D P V S Q L G L N
TCAGACAGCGTTCTTGGCTACAGCATTGGAGAAATCAAGCGCAGCAACACTTCCGAAACACCGGAGCTGC 630
open reading frame
S D S V L G Y S I G E I K R S N T S E T P E L
TTCCTTGTTGGCTACTGCTTGGAGAGAACACGACCATCTCAGTGACTGTGAAAGGGCTCGCAGAAAACAG 700
open reading frame
L P C G Y L V G E N T T I S V T V K G L A E N S
CCTGGACTCACTGCTGTTTGAAGTCTTATTCCTCAAGCCATCTGCAGCGCTACGTCTCCCTCCTGATA 770
open reading frame
L D S L V F E S L I P K P I L O R Y V S L L :
GAGCACCGTCGGATCATCTCTCTGGCCCCAGCGGCACCTGGGAAAACCTACCTGGCCAACCGGCTGTC 840
open reading frame
E H R R : I L S G P S G T G K T Y L A N R L S

FIG. 11a CONTINUED.

AGTATATAGTGCTTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGCCACCTTTAACGTGGACCATAA 910
open reading frame
E Y I V L R E G R E L T D G V I A T F N V D H K
GTCCAGCAAGGAATTGCGCCAGTACCTGTCCAACCTTGCTGACCAGTGCAACAGTGAGAACAATGCTGTG 980
open reading frame
S S K E L R Q Y L S N L A D Q C N S E N N A V
GACATGCCCCCTCGTCATCATCCTGGACAACCTACACCACGTGAGCTCTCTGGGCGAGATCTTCAATGGGC 1050
open reading frame
D M P L Y I I L D N L H H V S S L G E I F N G
TGCTCAACTGCAAGTACCACAAATGCCCTTACATAATTGGCACAATGAACCAGGCTACCTATCTCCCCCT 1120
open reading frame
L L N C K Y H K C P Y I I G T M N Q A T Y L P F
TTATACTAATAATCTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCTGCGGCCGC 1190
open reading frame LINKER-vector
Y T N N L I K K K K K K K K K K K K K K F C G R

FIG. 11b.

		20	20	30	40	50
1 1... 50 UNC53_huma	AAGCTTGGCA CGAGGCCTCG TGGCAAGCTG AGACCGTCAT GCAGCTCCGA					
5 1... 48	A W H E A S C Q A E T V M Q L R					
	60 LINKER > 70 ORF (START)					
1 51... 100 UNC53_huma	AATGAGTTAA GAGACAGGA GATGAAGCTG ACAGATATCC GCTTAGAAGC					
5 49... 98	N E L R D K E M K L T D I R L E A					
	110 120 130 140 150					
1 101... 150 UNC53_huma	TCTCAGTCTT GCCCACCAGC TGGACCAGCT CCGGGAGGCC ATGAACAGGA					
5 99... 148	L S S A H Q L D Q L R E A M N R					
	160 170 180 190 200					
1 151... 200 UNC53_huma	TGCAGAGTGA AATAGAGAAG CTGAAAGCTG AGAATGATCG GCTGAAGTCA					
5 149... 198	M Q S E I E K L K A E N D R L K S					
	210 220 230 240 250					
1 201... 250 UNC53_huma	GAGTCTCAAG GCAGTGGCTG CAGCCGGGCT CTTTCCAAG TGTCCATCTC					
5 199... 248	E S Q G S G C S R A P S Q V S I S					
	260 270 280 290 300					
1 251... 300 UNC53_huma	TGCCTCCCGG AGGCAGTCCA TGGGCTCTC CCAGCACAGC TTGAACCTCA					
5 249... 298	A S P R Q S M G L S Q H S L N L					
	310 320 330 340 350					
1 301... 350 UNC53_huma	CTGAGTCAAC CAGCCTGGAC ATGTGCTGG ATGACACTGG TGAATGCTCG					
5 299... 348	T E S T S L D M L L D D T G E C S					
	360 370 380 390 400					
1 351... 400 UNC53_huma	GCTCGGAAGG AAGGAGGCAG GCATGTTAAG ATAGTTGTCA GCTTCAGGA					
5 349... 398	A R K E G G R H V K I V V S F Q E					
	410 420 430 440 450					
1 401... 450 UNC53_huma	GGAAATGAAG TGAAGGAGG ATTCCAGACC ACACCTCTTT CTTATTGGCT					
5 399... 448	E M K W R E D S R P H L F L I G					
	460 470 480 490 500					
1 451... 500 UNC53_huma	GCATTGAGT TAGTGGCAAG ACGAAGTGGG ATGTGCTCGA TGGGGTGGTT					
5 449... 498	C I G V S G K T K W D V L D G V V					
	510 520 530 540 550					
1 501... 550 UNC53_huma	AGACGGCTGT TCAAAGATA CATCATTCAT GTGACCCAG TGAGTCAGCT					
5 499... 548	R R L F K E Y I I H V D P V S Q L					
	560 570 580 590 600					
1 551... 600 UNC53_huma	AGGCTTGAAT TCAGACAGCG TTCTGGCTA CAGCATTGGA GAAATCAAGC					
5 549... 598	G L N S D S V L G Y S I G E I R					
	610 620 630 640 650					
1 601... 650 UNC53_huma	GCAGGAACAC TTCCGAACA CCGAGCTGC TTCCTGTGG CTATCTGGTT					
5 599... 648	R S H T S E T P E L L P C G Y L V					
	660 670 680 690 700					
1 651... 700 UNC53_huma	GGAGAGAACA CGACCATCTC AGTGAATCTG AAAGGGCTCG CAGAAAACAG					
5 649... 698	G E N T T I S V T V K G L A E N S					

FIG. 11b CONTINUED.

	715	720	730	740	750
701..750 UNC53_huma	CCTGGACTCA	CTGGTGTTC	AGTCCTTGAT	TCCCAAGCCC	ATCCTGCAGC
5 699..748	L D S L V F E S L I P K P I L Q				
	765	770	780	790	800
751..900 UNC53_huma	GCTACGCTCTC	CCTCCTGATA	GAGCACCCTC	GGATCATTCT	CTCTGGCCCC
5 749..798	R Y V S L L I E H R R I I L S G P				
	810	820	830	840	850
801..850 UNC53_huma	AGCGGCACTG	GGAAAACCTA	CCTGGCCAC	CGGCTGTCTG	AGTATATAGT
5 799..848	S G T G R T Y L A N R L S E Y I V				
	860	870	880	890	900
851..900 UNC53_huma	GCTTCGAGAG	GGACGGGAGT	TGACAGACGG	GTTATCGCC	ACCTTTAACC
5 849..898	L R E G R E L T D G V I A T F N				
	910	920	930	940	950
901..950 UNC53_huma	TGGACCATAA	GTCCAGCAAG	GAATTGGCC	AGTACCTGTC	CAACCTTGCT
5 899..948	V D H K S S K E L R Q Y L S N L A				
	960	970	980	990	1000
951..1000 UNC53_huma	GACCACTGCA	ACAGTGAGAA	CAATGCTGTG	GACATGCCCC	TGCTCATCAT
5 949..998	D Q C N S E N N A V D M P L V I I				
	1010	1020	1030	1040	1050
1001..1050 UNC53_huma	CCTGGACAAC	GTACACCAGG	TGACCTCTCT	GGGCGAGATC	TTCAATGGCC
5 999..1048	L D N L H H V S S L G E I F N G				
	1060 PRIMER	1070	1080	1090	1100
1051..1100 UNC53_huma	TCTCAACTG	GAAGTACCAC	AAATGCCCTT	ACATAATTGG	CACAATGAAC
5 1049..1098	L L N C K Y H K C P Y I I G T M N				
	1110	1120	1130	1140	1150
1101..1150 UNC53_huma	CAGGCTACCT	CTTCGACTCC	CAACCTGCAG	CTTCACCATA	ACTTCAGATG
5 1099..1148	Q A T S S T P N L Q L H H N F R W				
	1160	1170	1180	1190	1200
1151..1200 UNC53_huma	GCTGCTTTGT	GCCAAACCACA	CGGAGCCTGT	GAAGGGTTTC	CTGGCCCGAT
5 1149..1198	V L C A N H T E P V K G F L G R				
	1210	1220	1230	1240	1250
1201..1250 UNC53_huma	TCTCGAGGAG	GAAGCTCATG	GAAACAGATA	TCAGTGGGCG	GCTGCGCAAT
5 1199..1248	F L R R K L M E T E I S G R V R N				
	1260	1270	1280	1290	1300
1251..1300 UNC53_huma	ATCGAGCTGG	TAAAAATCAT	TGACTGGATT	CCCAAGGTCT	GGCATCACCT
5 1249..1298	M I L V X I I D W I P K V W H H L				
	1310	1320	1330	1340	1350
1301..1350 UNC53_huma	CAACGCGCTC	CTGAGGCTC	ACAGTTCCTC	AGACGTGAC	ATCGCCCCC
5 1299..1348	N R F L E A H S S S D V T I G P				
	1360	1370	1380	1390	1400
1351..1400 UNC53_huma	GCTGCTTCCT	GTGATGCCCC	ATGATGTGCG	ACGGCTCGAG	AGTGTGGTTC
5 1349..1398	R L F L S C P I D V D G S R V W F				

FIG. 11b CONTINUED.

		1410	1420	1430	1440	1450
1 1401..1450 UNC53_hu	ACCGACTGT GGAACATATC CATTATCCCC TATCTCCTGG AAGCCGTCAG					
5 1399..1448	T D L W N Y S I I F Y L L E A V R					
		1460	1470	1480	1490	1500
1 1451..1500 UNC53_hu	AGAAGGACTC CAGCTCTATG GAAGCGCGCC CCGCTGGGAG GATCCTGCCA					
5 1449..1498	E G L Q L Y G R R A F W E D P A					
		1510	1520	1530	1540	1550
1 1501..1550 UNC53_hu	AGTGGGTGAT GGACACATAT CCGTGGGCAG CCAGCCCAACA ACAGCAGGAG					
5 1499..1548	K W V M D T Y P W A A S P Q Q H E					
		1560	1570	1580	1590	1600
1 1551..1600 UNC53_hu	TGGCCTCCCC TCGTGCAGTT ACGGCCTGAG GATGTCGGCT TCGACGGCTA					
5 1549..1598	W P P L L Q L R P E D V G F D G Y					
		1610	1620	1630	1640	1650
1 1601..1650 UNC53_hu	CTCCATGGCT CGGGAGGGAT CGACAAGCAA GCGATCCCC CCGAGTATG					
5 1599..1648	S X P R E G S T S K Q M P P S D					
		1660	1670	1680	1690	1700
1 1651..1700 UNC53_hu	CTGAAGGTGA CCGCTGATG AACATGCTGA TGAGGCTGCA GGAAGGAGCC					
5 1649..1698	A E G D P L M N M L M R L Q E A A					
		1710	1720	1730	1740	1750
1 1701..1750 UNC53_hu	AACACTCCA GCGCCAGAG CTATGACAGC GACTCCACA GCAACAGCCA					
5 1699..1748	N Y S S P Q S Y D S D S N S N S H					
		1760	1770	1780	1790	1800
1 1751..1800 UNC53_hu	TCAGATGAC ATCTTGGACT CCTCTTTGGA GTCCACTCTG TGCAGGGGC					
5 1749..1790	H D D I L D S S L E S T L					
		1810	1820	1830	1840	1850
1 1801..1850 UNC53_hu	CCGAGGCCA GCGCCCTCCT CTCTCTCTCA CCGATTCCA CCGCATCCC					
5 ----	<==					
		1860	1870	1880	1890	1900
1 1851..1900 UNC53_hu	CAGATCAGCC TGAAGATGAC TCCCTGAGCC AGCCCCAGCC ACAGCCTTAG					
5 ----	<==					
		1910	1920	1930	1940	1950
1 1901..1941 UNC53_hu	AGCTCGGGA ACACCGAGAC CCCCCTCCTT CAGCCTGAC					
5 ----	<==					

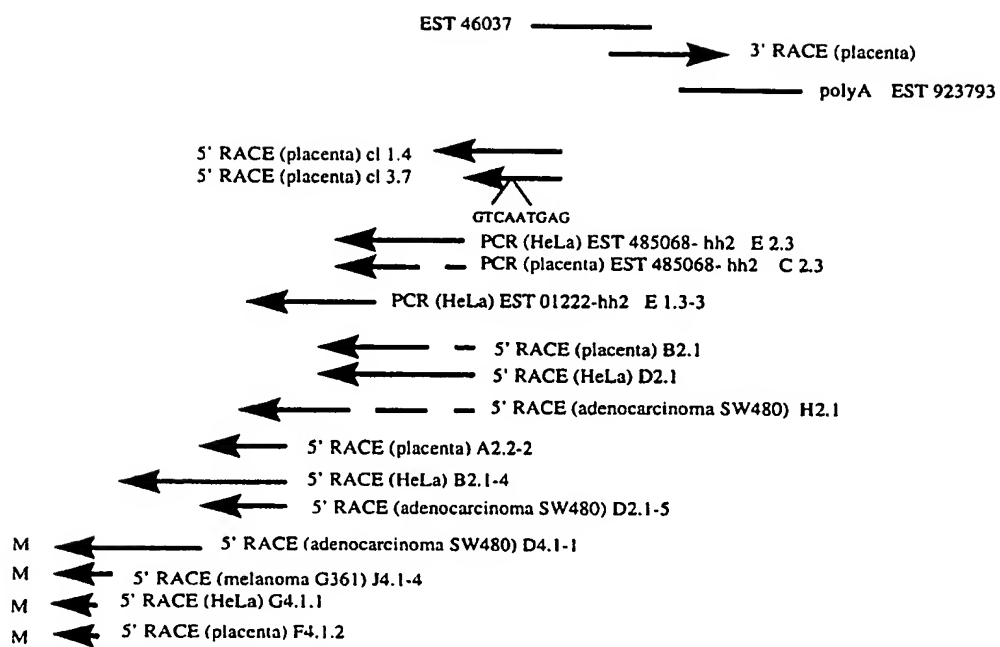


Figure 11c

[illegible][illegible]

FIGURE 11d

-1 I P S E A H W S A P B S N S T W G T N A S S S S A V S
 3841 ACACUAGAT CACUCUAGAT CCACUCUAGAT MACUCUAGAT CUCUCUCUAGAT ACACUCUAGAT ACUCUAGAT ACUCUAGAT ACUCUAGAT
 TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT TOCCUCUAGAT

-1 D G L Q G F O S V S S L M T S C E S S I D I S L S S S G G
 3921 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 V P S H S T G L I N A S K D S L T P F V R T U S
 4001 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 V K T L S E S P L S S P A A S P K F C R S S T L P R
 4081 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 O D S D P H L D R N T L P K K L R Y T P T S O L R
 4161 ACACUAGAT CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 T Q E D A K E W L R S H S A G L O D T A A H S P F S
 4281 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 G S S V T S P S G T R F N F S Q L A S P T V T Q
 4321 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 S L S N P T M L R T H S L S N A D G O Y D P Y T D S
 4401 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 R F R N S S M S L D F E S R T M S R S G S P R D G F E
 4481 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 F V H G S L S L V S S T L S V Y S T P F E F C Q S
 4561 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 I R K L R L R L D A S Q E K V S A L T F O L T A N A
 4641 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 H L V A A F E Q S L G N M T I R L O S L T M T A E Q K
 4721 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 D S E L N E L R K T I E L L K K Q N A A Q A A I N
 4801 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 V I N T P E L N C K G N G T A Q S A D L R I R R O H
 4881 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 S S D S V S S I N S A T S H S S V G S N I E S D S K K
 4961 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

-1 E K R K H M L R S S P K O A F G K K S P K S A S S
 5041 CACUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT
 CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT CUCUCUAGAT

[illegible]

11/24/97 02:33PM

Fig 11c

EST 223793

Page 1

```
+1 G T R V T I G P R L F L S C P I D V D G S R V W F T D
1 GGCACGAGGG TTACCATCGG CCCCCGGGCTC TTCCTGTCAT GCCCCATCGA TGTGGACGGC TCGAGAGTGT GGTTCACCGA
CCGTCTCCC AATGGTAGCC GGGGGCCGAG AAGGACAGTA CGGGGTAGCT ACACCTGCCG AGCTCTCACA CCAAGTGGCT
.....
+1 L W N Y S I I P Y L L E A V R E G L Q L Y G R R A P
81 CTTGTGGAAC TATTCATTA TCCCCTATCT CCTGGAAGCC GTCAGAGAAG GACTCCAGCT CTATGGAAGG CGCGCCCCCT
GAACACCTTG ATAAGGTAAT AGGGGATAGA GGACCTTCGG CAGTCTCTTC CTGAGGTGGA GATACCTTCC GCGCGGGGGA
.....
+1 W E D P A K W V M D T Y P W A A S P Q Q H E W P P L L
NcoI
-----
161 GGGAGGATCC TGCCAAGTGG GTGATGGACA CATATCCATG GGCAGCCAGC CCACAACAGC ACGAGTGGCC TCCCCTGCTG
CCCTCCTAGG ACGGTTTCACC CACTACCTGT GTATAGGTAC CCGTCGGTCG GGTGTTGTCT TGCTCACCAG AGGGGACGAC
.....
+1 Q L R P E D V G F D G Y S M P R E G S T S K Q M P P S
241 CAGTTACGGC CTGAGGATGT CGGCTTCGAC GGCTACTCCA TGCCCTCGGA GGGATCGACA AGCAAGCAGA TGCCCCCAG
GTCAATGCCG GACTCCTACA GCCGAAGCTG CCGATGAGGT ACGGAGCCCT CCCTAGCTGT TCGTTCGTCT ACGGGGGGTC
.....
+1 D A E G D P L M N M L M R L Q E A A N Y S S P Q S Y
321 TGATGCTGAA GGTGACCCGC TGATGAACAT GGTGATGAGG CTGCAGGAGG CAGCCAACTA CTCAGCCCC CAGAGCTATG
ACTACGACTT CCACTGGGCG ACTACTTGTA CGACTACTCC GAGTCTCTCC GTCGGTTGAT GAGGTCGGGG GTCTCGATAC
.....
+1 D S D S N S N S H H E D I L D S S L E S T L * Q G P G
401 ACAGCGACTC CAACAGCAAC AGCCATCAGC AAGACATCTT GGAATCCTCT TTGGAGTCCA CTCTGTGACA GGGGCCCCGA
TGTCGCTGAG GTTGTCTGTG TCGGTAGTGC TTCTGTAGAA CCTGAGGAGA AACCTCAGGT GAGACACTGT CCCCAGGCGT
.....
+1 A Q R P P L L L T A F H L H P P H H P E D D F L S Q P
481 GCCCAGCGCC CTCCTCTTCT CCTCACCACA TTCCACCTGC ATCCCCCACA TCACCCTGAA GATGACTTCC TGAGCCAGCC
CGGGTCGGCG GAGGAGAAGA GGAGTGGCGT AAGGTGAGC TAGGGGGTGT AGTGGGACTT CTACTGAAGG ACTCGGTCCG
.....
+1 P A T A L E L R E H R D P P S P S L D L G A G I P G
561 CCCAGCCACA GCCTTAGAGC TCGGGGAACA CCGAGACCCC CCGTCCTTCA GCCTCGACCT GGTGTCAGGC ATCCCGGGCC
GGGTCTGTGT CGGAATCTCG ACGCCCTTGT GGCTCTGGGG GGCAGGAAGT CGGAGCTGGA CCCACGTCCG TAGGGCCCCG
.....
+1 Q L P A D R F L P Q R E L H Y L L L Y F N Y C F A L L
641 AGCTGCCTGC GGACCGCTTC CTTCCACAGC GAGAATGCA CTACCTTCTG TTGTACTTTA ATTATTGTTT TGCCTTGTTG
TCGACGGACG CCTGGCGAAG GAAGGTGTCG CTCTTGACGT GATGGAAGAC AACATGAAAT TAATAACAAA ACGGAACAAC
.....
+1 L * P P * D T E D T S R E R I I A V E M K K K K K K K
721 CTGTGACCTC CTAAGACAC TGAAGATACT TCTCGGAAAA GGATCATCGC CGTTGAAATG AAAAAAAAAA AAAAAAAAAA
GAACTGGAG GGATTCTGTG ACTTCTATGA AGAGCCCTTT CCTAGTAGCG GCAACTTTAC TTTTTTTTTT TTTTTTTTTT
.....
+1 K K K K K N E G G R K L
801 AAAAAAAAAA AAAAAAACG AAGGCGGCGG CAAGCTT
TTTTTTTTTT TTTTTTTTGC TTCCGCGGCG GTTCGAA
```

11/24/97 02:40PM

hh2UNC53_5'Race_cl1.4_CON_REV.seq

Page 1

+1 I P L T I G L E R P P R Q V P H L D R N T L P K K G L
1 ATCCCACTCA CTATAGGGCT CGAGCGGCGG CCCAGGCAGG TCCCGCACCT TGATAGGAAC ACTTTGCCTA AGAAAAGGACT
TAGGGTGAGT GATATCCCGA GCTCGCCGCG GGGTCCGTCC AGGGCGTGGA ACTATCCTTG TGAAACGGAT TCTTTCTCTGA
.....
+1 R Y T P T S Q L R T Q E D A K E W L R S H S A G G L
81 CAGGTATACT CCCACCTCCC AGCTTCGCAC GCAAGAAGAT GCAAAAGAAT GGTTACGGTC CCATTCTGCA GGAGGCCTTC
GTCCATATGA GGGTGGAGGG TCGAAGCGTG CGTTCTTCTA CGTTTTCTTA CCAATGCCAG GGTAAGACGT CCTCCGGAAG
.....
+1 Q D T A A N S P F S S G S S V T S P S G T R F N F S Q
161 AGGACACCGC TGCCAATTCC CCCTTTTCTT CTGGCTCCAG CGTGACTTCT CCCTCCGGA CAAGATTCAA CTTTTCCTCAG
TCTGTGGCG ACGGTTAAGG GGGAAAAGGA GACCGAGGTC GCACTGAAGA GGGAGGCCTT GTTCTAAGTT GAAAAGGGTC
.....
+1 L A S P T T V T Q M S L S N P T M L R T H S L S N A D
241 CTTGCGAGTC CCACCACTGT CACCCAGATG AGCTTGTCCTA ACCCGACCAT GCTGAGGACT CACAGCCTCT CCAATGCTGA
GAACGCTCAG GGTGGTGACA GTGGGTCTAC TCGAACAGGT TGGGCTGGTA CGACTCCTGA GTGTCCGGAGA GGTTCAGACT
.....
+1 G Q Y D P Y T D S R F R N S S M S L D E K S R T M S
321 TGGGCAGTAT GATCCATACA CTGACAGCGG CTTCCGGAAT AGCTCCATGT CCCTGGATGA GAAGAGCAGA ACCATGAGCC
ACCCGTCATA CTAGGTATGT GACTGTGCGC GAAGGCCCTA TCGAGGTACA GGGACCTACT CTTCTCGTCT TGGTACTCGG
.....
+1 R S G S F R D G F E E V H G S S L S L V S S T L S V Y
401 GTTCAGGCTC ATTCCGGGAT GGGTTTGAAG AAGTTCATGG ATCCTCACTC TCCCTGGTTT CCAGCACATT GTCAGTTTAT
CAAGTCCGAG TAAGGCCCTA CCCAAACTTC TTCAAGTACC TAGGAGTGAG AGGGACCAAA GGTCGTGTAA CAGTCAAATA
.....
+1 S T P E E K C Q S E I R K L R R E L D A S Q E K V S A
481 TCTACACAG AAGAAAAATG CCAGTCAGAG ATTCGCAAGC TGCGGCGGGA ACTGGATGCC TCCCAGGAGA AAGTTTCAGC
AGATGTGGTC TTCTTTTAC GGTCACTCTC TAAGCGTTTC ACGCCGCCCT TGACCTACGG AGGGTCTCTT TTCAAAGTCG
.....
+1 L T T Q L T A N A H L V A A P E Q S L G N M T I R L
561 LTTGACCACC CAGCTGACAG CAAATGCTCA CCTGTGGCT GCCTTTGAAC AGAGTCTGG TAACATGACA ATCAGGCTCC
AAACTGGTGG GTCGACTGTC GTTTACGAGT GGAACACCGA CGGAAACTTG TCTCAGAACC ATTGTACTGT TAGTCCGAGG
.....
+1 Q S L T M T A E Q K D S E L N E L R K T I E L L K K Q
641 AGAGTCTGAC CATGACAGCT GAGCAGAAGG ATTCGAACT GAATGAGTTA AGAAAAACCA TTGAGCTGCT AAAGAAACAG
TCTCAGACTG GTACTGTCGA CTCGTCTTCC TAAGTCTTGA CTTACTCAAT TCTTTTGGT AACTCGACGA TTTCTTTGTC
.....
+1 N A A A Q A A I N G V I N T P E L N C K G N G T A Q S
721 AACGCAGCTG CCCAGGCTGC CATTAAATGGA GTAATTAACA CACCTGAGCT CAACTGCAAA GGAAACGGCA CTGCCAGCTC
TTGCGTCGAC GGGTCCGACG GTAATTACCT CATTAAATGT GTGGACTCGA GTTGACGTTT CCTTGGCGT GACGGGTCAG
.....
+1 A D L R I R R Q H S S D S V S S I N S A T S H S S V
801 TGCAGACCTC CGCATCCGCA GGCAGCACTC CTCAGACAGC GTCTCCAGCA TCAACAGTGC CACCAGCCAC TCCAGTGTGG
ACGTCTGGAG GCGTAGGCGT CCGTCTGTAG GAGTCTGTCT CAGAGGTCGT AGTTGTACAG GTGGTCCGTC AGGTACACCC
.....
+1 G S N I E S D S K K K K R K N W L R S S F K Q A F G K
881 GCAGCAACAT AGAGAGTGAC TCAAAGAAGA AGAAGAGGAA GAACTGGTTA CGCAGCTCCT TCAAGCAAGC TTTCCGGGAAG
CGTCGTTGTA TCTCTCACTG AGTTTCTTCT TCTTCTCCTT CTTGACCAAT GGTCTGAGGA AGTTCTGTTC AAAGCCCTTC
.....
+1 K K S P K S A S S H S D I E E T D S S L P S S P K L
961 AAGAAGTCCC CAAAATCTGC GTCTCTCAT TCAGATATTG AGGAGACGAC GGATTCTTCT TTGCCTTCT CACCAAAGTT
TTCTTCAGGG GTTTTAGACG CAGGAGAGTA AGTCTATAAC TCCTCTGCTG CTAAGAAGA AACGGAAGGA GTGGTTTCAA
.....
+1 P H N G S T G S T P L L R N S H S N S L I S E C M D
1041 ACCGCACAAT GGGTCCACAG GTTCCACCCC ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTCGGA TGCATGGATA
TGGCGTGTTA CCCAGGTGTC CAAGGTGGGG TGACGACTCC TTAAGAGTGA GGTGAGAGA TTAAGGCTT ACGTACCTAT
.....
+1 S E A E T V M Q L R N E L R D K E M K L T D I R
1121 GTGAAGCTGA GACCGTCATG CAGCTCCGAA ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CGGATATCCG
CACTTCGACT CTGGCAGTAC GTCGAGGCTT TACTCAATTC TCTGTTCCTC TACTTCGACT GCCTATAGGC G

11/24/97 02:46PM

hh2UNC53_5'Race_cl3.7_CON.seq

Page 1

+2 E F E L G T L T I G L E R P P G Q V R D G F E E V H
1 CGAATTCGAG CTCGGTACAC TCACTATAGG GCTCGAGCGG CCGCCCGGGC AGGTCCGGGA TGGGTTTGAA GAAAGTTCATG
GCTTAAGCTC GAGCCATGTG AGTGATATCC CGAGCTCGCC GCGGGGCCCC TCCAGGCCCT ACCCAAACCT CTCAAGTAC

+2 G S S L S L V S S T S S V Y S T P E E K C Q S E I R K
81 GATCCTCACT CTCCTTGTT TCCAGCACAT CGTCAGTTTA TTCTACACCA GAAGAAAAT GCCAGTCAGA GATTGCAAG
CTAGGAGTGA GAGGAACCA AGGTCGTGTA GCAGTCAAAT AAGATGTGGT CTTCTTTTFA CGGTCACTCT CTAAGCGTTC

+2 L R R E L D A S Q E K V S A L T T Q L T A N A H L V A
161 CTGCGGCGGG AACTGGATGC CTCCAGGAG AAAGTTTCAG CTTTGACCAC CCAGCTGACA GCAAATGCTC ACCTTGTTGGC
GACGCGCCCC TTGACCTACG GAGGTCCTC TTTCAAAGTC GAAACTGGTG GGTGACTGT CGTTTACGAG TGAACACCG

+2 A F E Q S L G N M T I R L Q S L T M T A E Q K D S E
241 AGCCTTTGAA CAGAGTCTTG GTAACATGAC AATCAGGCTC CAGAGTCTGA CCATGACAGC TGAGCAGAAG GACTCAGAAC
TCGGAACCTT GTCTCAGAAC CATGTACTG TTAGTCCGAG GTCTCAGACT GGTACTGTG ACTCGTCTTC CTGAGTCTTG

+2 L N E L R K T I E L L K K Q N A A A Q A A I N G V I N
321 TGAATGAGTT AAGAAAAACC ATTGAGCTGC TAAAGAAACA GAACGAGCT GCCCAGGCTG CCATTAATGG AGTAATTAAC
ACTTACTCAA TTCTTTTGG TAACTCGACG ATTTCTTTGT CTGCGTGA GGGTCCGAC GGTAATTACC TCATTAATTG

+2 T P E L N C K G N G T A Q S A D L R I R R Q H S S D S
401 ACACCTGAGC TCAACTGCAA AGGAAACGGC ACTGCCAGT CTGCAGACCT CCGCATCCGC AGGCAGCACT CCTCAGACAG
TGTGGACTCG AGTTGACGTT TCCTTTGCCG TGACGGGTCA GACGTCTGGA GGGTAGGCG TCCGTCTGTA GGAGTCTGTC

+2 V S S I N S A T S H S S V G S N I E S D S K K K K R
481 CGTCTCCAGC ATCAACAGTG CCACCAGCCA CTCCAGCGTG GGCAGCAACA TAGAGAGTGA CTCAAAGAAG AAGAAGCGGA
GCAGAGGTCG TAGTTGTCAC GGTGGTCGGT GAGGTGCGAC CCGTCGTTGT ATCTCTCACT GAGTTCTTC TTCTTCGCTT

+2 K N W V N E L R S S F K Q A F G K K K S P K S A S S H
561 AGAAGTGGT CAATGAGTTA CGCAGCTCCT TCAAGCAAGC TTTGGGAAG AAGAAGTCCC CAAATCTGC GTCTCTCAT
TCTTGACCCA GTTACTCAAT GCGTCGAGGA AGTTCGTTG AAAGCCCTTC TTCTCAGG GTTTAGACG CAGGAGAGTA

+2 S D I E E M T D S S L P S S P K L P H N G S T G S T P
641 TCAGATATTG AGGAGATGAC GGATTCTTCT TTGCCCTTCT CACCAAAGTT ACCGCACAAT GGGTCCACAG GTTCCACCCC
AGTCTATAAC TCCTCTACTG CCTAAGAAGA AACGGAAGGA GTGGTTTCAA TGGCGTGTTA CCCAGGTGTC CAAGTGGGG

+2 L L R N S H S N S L I S E C M D S E A E T V M Q L R
721 ACTGCTGAGG AATTCTCACT CCAACTCTCT AATTTCAGAA TGCATGGATA GTGAAGCTGA GACCGTCATG CAGCTCCGAA
TGACGACTCC TTAAGAGTGA GGTGAGAGA TTAAGTCTT ACGTACCTAT CACTTCGACT CTGGCAGTAC GTGAGGCTT

+2 N E L R D K E M K L T D
801 ATGAGTTAAG AGACAAGGAG ATGAAGCTGA CGGATAT
TACTCAATTC TCTGTTCTC TACTTCGACT GCCTATA

[illegible]

Page 1

9

[illegible]

[illegible]

[illegible]

1 R N T L P K K G L R V
1281 TAGAGCACT TTACCTAGCA AGCAGTAC GCTMA
ATCCTGTCA AGCAGATCT TTCTAGAC CATT
.....

[illegible]

1281 CAGAGAGCC TCTCTCTCC CTCCTGTCA CCTCAATAT CTCCTGATC CAGAAAGCA CAGTACCTCCCTC
 GCTTCTCGG AGAGACAGG GAGACATC CCGATCTTA CAGCTCTTA CAGAAAGAT CTCCTCTTC GTTACTATTTG
 1361 CAGCTTATC GAGAAATTC CAGTAAAG GAGTACATC
 GTGACACTT CCGTCTTAA CCGATCTTC CCGTAAACA AA

21/26/97 03:46PM

b725' raceH2-1.COM

Page 1

[illegible]

11/24/97 03:46PM

hh25' raceH2-1_COM

Page 2

1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
19

11/24/97 04:11PM

D2.1-5 (In)DUNC3 5 PAGE 02/04/97)

Page 1

*) S L * G S S G R P G R C S S T P T H C

1 GATITICA GATITICIT TATITICIT AGATITICAG GATITICAGG GATITICAGG GATITICAGG TATITICAGG TATITICAGG
GATITICAGG GATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG
TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG TATITICAGG

*) S T S A I P O P G A A T R P M R S K S L S V K H S A
81 AGATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
TATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) T V S M L S V K P P G P E A P H P T P E A N K P A P
161 GATITICIT AGATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT
GATITICAGG TATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG

*) R O K S M L E K L K L F N S K G G S K A G E G P O S
241 AGATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
TATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) R D T S C E R L E T L P S P E E S E L E A A S R M L
321 GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) T T V G P A S S S P K I A L K G I A O R T P S R A L
401 GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) T N K K S S L K G N E K E K E K O O N E K D E K S K
481 GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) O L A K R A S V T E R L D L K E E P R E D P S G A A V
561 GATITICIT AGATITICIT TATITICIT AGATITICIT TATITICIT AGATITICIT TATITICIT AGATITICIT TATITICIT
GATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG

*) P E M P K R S S K I A S P I P K G G L N S A K K E
641 GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) P H A P S H S G I P K P G M R S N P G K S P S A P A P
721 GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT GATITICIT
GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG GATITICAGG

*) S K E G E R S R S G K L S S G L P O O K P O L D G R H
801 TATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG
AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT

*) S S S S S L A S S E G K G P O G T T L N H S I S S
881 GATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT TATITICIT
GATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG

*) Q T V S G S V G T T O T T G S N T V S V O L P O P O Q
961 AGATITICAGG TATITICIT AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG
TATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT

*) Q V H P M T A T V A P P L V R S O T D T E G N V T A
1041 GATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG
AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT

*) E S S S T G V S V E P S H P T K T G O P A L E E L T
1121 GATITICAGG TATITICIT AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG
AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT

11/24/97 04:11PM

D2.1-5 (In)DUNC3 5 PAGE 02/04/97)

Page 2

*) G E D P E A R R L R T V K H I A O L R G H L S E T V S
1201 AGATITICIT TATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT
AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG

*) S L R G T O V T K P H
1281 AGATITICIT GATITICIT TATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT AGATITICIT
AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG AGATITICAGG

[illegible]

Fig. 11f.

Page 1

UNRS31h2_5'race_D4.1.1_CON.seq

MSO:FO 66/PC-7

[illegible]

P

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

1 CTATACAGU TATATACCA AGGATTTAT CTATACACC GAGACAGC CTATATACC AGTTATCCA GATTCAGAT
GAGATTTCT AGATCTCT TCCAMAHIA GAGTCTCC GCTCTCTCC GAGCAACCT TCAATAGCT GAGAGGTAA
..... P P P R D T G
81 GCGCGGAGA AGCAGCGAG TCGCTACGA GAGACAGC ATCGCTGCT TTCTCTTCC TCTTCTTCA GCGACAGAG
CGGAGCTCT TCGTCTCTC AGCGATCT CTCTCTTCC TCGACAGC AGCAGAGCT GCGAGAGCT CTCTCTCTC
.....
82 R O T L F A E T S T A F I I T K P L P C S D N S F O
161 CAGACAGA CTCTCTCTC AGCTCTCAG AGCTCTCTT ATTCAGAAC CTCTCTCC CTCTCTCC GATTCAGAT GATTCAGC
GCTCTCTCT GCGAGAGCT TCGAGAGCT TCGAGAGTA TATCTTTTC GAGACAGC GATCTCTTA CTAGAGCTC
.....
83 Q M A L P R A S V E H L E O L D F I E C L P C T
241 AATCTAGC AGTCTCTC AGAGCTCTC TCGAGCTCT AGATACAA CTGAGCTTTA TTAGCTCTT ACTATCAGC
TTATCTCTC TATATAGC TCTCTAGC ATCTCTGA TCTATTTT GACTCTAAT ACTATACAA TATATCTG
.....
84 K P W A K H F I C R L F V F T A B P V G R Y N Y P H S
321 AGCTCTGAG CTATAGCTT CATCTCAG CTCTCTCTT TTAGCGAAA CCGATAGCT AGTATAGCT ATCCGACTC
TTAGCGAGC GTATCTGAG GTAGAGCTCC GAGAGCGAGA ATCTCTCTT GCGTCTATCA TCGATTTTGA TTAGCTTAG
.....
85 A D A E T E A Q S V L V A K Q A H Q E A R K W P H L
401 TCAATATCA GAGAGAGAG CAGAGAGCT TTCTATAGT AAAAAGCTC ACTAGAGCTC TACAGAGCTC GCGAGCTAG
AGCTATAGT CTCTCTCTC CTCTCTCTC AGAGCTATCA TTCTCTAG TATCTCTCC ATCTCTCACC GCTCTCTC
.....
86 A G P P D S T N C L P L L C C H Q E C D S K F F L P S
461 CTGAGCTCTC TACTCTCAG ACTCTCTCC CTCTCTCTC TTGCTATGAA GATCTCTCTT CCACTCTTT CTCTCTCT
GAGCTGAGCTC ACTAGCTGCT TTAGCTAGC GAGAGAGAG AGCTATAGT CTATCTCTA GCTTCAMAA GAGAGAGAA
.....
87 G S H S G P T L L S H Q I Y T D W A N H Y L A K S G H
561 GATCTAGCT CTGCTCTC TCTCTCAG AACAGATCT AGAGAGCTC GCGATAGT TACTAGCA ACTCTCTC
CTATAGCTCA GAGCAAGCT AGAGAGCTC TTGCTATGAA TCTCTCTC GCGTATAGT ATGAGCTCT TTAGCTCT
.....
88 K R L I K D L Q E S R I P A H W R P L L V D P S S
641 CAGCTCTC ATAGAGATC TCGAGAGAA AGCGAGAT GCGAGACT GCGAGCTCT ACTAGCTAT CCGAGCTG
GCTCTCTCAG TATCTCTAG AGCTCTCTT TTGCTTTTAA GCTCTCTTAA CCGCGGAA TATCTCTA GCTCTGAGC
.....
89 V P K L G V I M V I V
721 TACGAGCT TCGCTATC ATCTCTAG TGT
ATCTCTCA ATCTCTAG TACTCTATC ACA

R

1 RRRAGAGCA GGGGGTAAA AAGGTTTTH AUAAGAGG GATCCCTTG NGGGGTHT TTTTNGCC CCNKGHTG
KGTCTTCT CCCTATTT TTRHAAAT TTTTTRCC CTRGGGAC NCCCAAAA AAAAGGCG GGGGCGAAC
81 TAAAGGGG TCCCTTNGG ATTTTHANG GAGGGGAG GAGGGGAG ACCCTTCT TCGGGGGG GTCAAGCG
ATTTTCCC AGGGAATCC TAAHAAATC CTTTCTGTC CTTCTCTG TCGAAGAA ACCCGCCG CACTTNGG
161 KCAACACT GGGGAGGAC KCGGAGCAI TTTTCTTT AUAAGGGA AATTGAGAG AAGGGGCG TTTGGTAA
RTGGTGGCA CCCCCTCG KCCCTGTH AUAAGGAAA TTTTCTCT TAAAGCTC TTTTCCGC AAACCAATC
241 TCGGGGAT TTTGGGNC ATNAGATG AGAGTAGG TGGATCCG CTTAGNGG GGHATGTC NGGGTCTCT
HACCTTCA AAGCCGAG TTAATGTA TGTCTATC ACCTAGAG GATTTGCC CCAATACAG NCCCAAGG
321 AGTTGCGG TTTTCTCG CAGGTGCA AGGGAAGG TTAAGAAI TCGATTAI CCGGTHAA TCAHACTT
TAAAGCC AUAAGAGG CTGAGGCT TCGGTTCC AATCTCTA AGGTHATT GAGCAATT ACTHTGAA
401 TGGTTTTT HTATTHAA ATATTHAA TAAHAAAT TTTCTTCC CAGAAAAA AAAAGAGA AATTAACA
ACTAAAAA HATHTATT TATHTATT ATHTHTTA AAGGAAGG GTTTTTTT TTTTCTCT TTHATHTG
481 HAAATATA HAAAAAAN GAAAAAGI AATCAAAA NAAAAAGG GCGTTAAI TTTAGCTC TAAAGGATG
HTATHTAT HTTTTCTH CTTTCTTH TTTAGTTTT TTTTTCGC CCGNACTA AATCTGAG ATCTGTAC
561 V R P H S L S R D R V V R V P R A P I V S . Y F G V
TGTGCGCT CAGAGCTCT CCGGAGAG AGTGTCCG GTCCCGGAG CCCCATTCT GTCTGTAC TTTGGGTC
ACAGCGGA GTTGGGCA GGGCTCTC TCACAGCG CAGGGGCTC GGGGTACA CAGACTATG AATCCGAG
641 H M A I D L Y C G L A C L W G K H E P R I Y T D W A N
AGATGCTAT AGATTTTAC TCGGTTTG CTTGTCTG GGAAGAAI GAGCCCGA TCTACAGA CTGGGCAI
TGTACGATA TTAAGATG ACCCAAGC CAGAGAGC CCGTTTCTA CTGGGCTT AGATCTCT GAGCCGTA
721 H Y L A K S G H K R I I K D F Q Q E S R I L O I S I T
CATACTAG CCAATCGG CCAAGCTT ATCTAGG ATTTGAGA AUAAGGCA ATTCTGAG TATCATAC
GTATGATC GTTTAGCC GTTTGCGA TATGATCC TAAAGTCT TTTTCTCT TAAAGCTCT AATGTAAG
801 X A A A B A C X . R P I R P I V S C I Q
ATGCGGCT GTGAGAGT GAGTAGAG GCGATCGC CCAATAGCA GTTGTATCA AT
TAAAGCGG GAGTCTGA CTAHATCT CCGTAGAG GATATCAT CAGATATG TA

Figure 12. Tblastn search of the EST division of Genbank with 680aa starting at the c-terminus of the alfa-actinin domain of Hu-UNC-53/2.

LOCUS AA418158 610 bp mRNA EST 19-MAY-1997
 DEFINITION zv97dl2.r1 Soares NhBMPu S1 Homo sapiens cDNA clone 767735 5'.
 ACCESSION AA418158
 NID g2079968
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
 Homo.
 REFERENCE 1 (bases 1 to 610)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
 Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 TITLE WashU-Merck EST Project 1997
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 492.
 FEATURES
 source Location/Qualifiers
 1..610
 /organism="Homo sapiens"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not I;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NbEM, pregnant uterus
 NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."
 /clone="767735"
 /clone_lib="Soares NhBMPu S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 mRNA <1..>610

```

                                /clone="5D16"
                                /clone_lib="Zebrafish ICRFzf1s"
                                /sex="mixed"
                                /tissue_type="pooled 26-somite embryos"
                                /lab_host="XL1-blue MRF"
                                complement(<1..>418)
mRNA
BASE COUNT      108 a      87 c      78 g      145 t
ORIGIN
    1 tttacatttt ttgaggaaga tgctaattggt ctattctgat tcaatgattt atgctaagct
   61 aagctaaaaat gtcctgtgca aatcctgaga tcagctgaat gaattaaaaa ttgggtaaaa
  121 ctcaactgtc taactctagg ggagttgtaa aatgggccta tttccctaaa aagtaatggt
  181 actttaagag catgatggtc caccagtttc actgtctaaa ttttgttatt ccataagcta
  241 atcttctctg ggcattttga cgattttaac actaacctgt gggtaatctg cgcccccggt
  301 aaactggaca tggtttcttc cagattctgt ctcagatcag caatgttctt cactgtacgc
  361 atccgtctag tttctggatc ttctctgag atctcctcca ggcactgttt ggcgggtct
//
gb|AA495042|AA495042 fa05f06.s1 Zebrafish ICRFzf1s Danio rerio cDNA
    clone 5D16 3'
    Length = 418

```

Minus Strand HSPs:

Score = 195 (87.9 bits), Expect = 9.9e-18, P = 9.9e-18
 Identities = 37/46 (80%), Positives = 42/46 (91%), Frame = -3

Query: 627 TGQPALEELTGEDPEARLRRTVKNIADLRQNL EETMSSLRGTVTH 672
 T + LEE++GEDPE RR+RTVKNIADLRQNL EETMSSLRGTV+TH
 Sbjct: 416 TAKQCLEEISGEDPETRRMRTVKNIADLRQNL EETMSSLRGTVITH 279

MOUSE 2

```

LOCUS      AA208994      527 bp      mRNA      EST      18-FEB-1997
DEFINITION mw75e12.r1 Soares mouse NML Mus musculus cDNA clone 676558 5'.
ACCESSION  AA208994
NID        gl807004
KEYWORDS   EST.
SOURCE     house mouse.
  ORGANISM Mus musculus
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Mus.
REFERENCE  1 (bases 1 to 527)
  AUTHORS  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE    The WashU-HHMI Mouse EST Project
  JOURNAL  Unpublished (1996)
COMMENT
  Contact: Marra M/Mouse EST Project
  WashU-HHMI Mouse EST Project
  Washington University School of MedicineP
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:416262
 Putative full length read
 vector to vector length is 535
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 478.

FEATURES
 source Location/Qualifiers
 1..527
 /organism="Mus musculus"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAAGTGGGAGCGGCCGCGAATCTTTTTTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."
 /clone="676558"
 /clone_lib="Soares mouse NML"
 /tissue_type="Liver"
 /lab_host="DH10B"
 mRNA
 <1..>527
 BASE COUNT 151 a 139 c 136 g 101 t
 ORIGIN
 1 tgtctctgga tgagaagagc cgaacaatga gtcggtcagg ctccttccgg gatggggttg
 61 aggaagtcca tggatcctcc ctgtccttgg ttccagcac atcctccatc tactccacgc
 121 cagaagaaaa atgccagtca gagattcgaa agctgaggcg agacgtggat gcctcccagg
 181 aaaagggtgc tgcgctgact acccagctga ctgcaaatgc tcaccttgtg gcagccttgc
 241 agcagagtct gggaaacatg accatcaggc tacagagttt aactatgacc gctgagcaga
 301 aggattcaga actgaacgag ttaagaaaaa ccatcgagct gctgaagaaa cagaatgcag
 361 ctgcccaggc tgccattaat ggagtgatta acacgccaga gctcaactgc aaaggaaatg
 421 gcagtgccag gctacagacc tacgcatccg cagcaacact cctccgacag tgtctccagt
 481 atcaatagcg ccaccagcca ctcaagtgtg ggcagcaaca tagagag

gb|AA208994|AA208994 mw75e12.r1 Soares mouse NML Mus musculus cDNA
 clone 676558 5'
 Length = 527

Plus Strand HSPs:

Score = 541 (243.9 bits), Expect = 2.3e-76, Sum P(2) = 2.3e-76
 Identities = 110/143 (76%), Positives = 114/143 (79%), Frame = +3

Query: 1511 SLDEKSRMSRSGSFRDGFEEVHGXXXXXXXXXXXXXXXXXPEEKQSEIRKLRLRELDASQE 1570
 SLDEKSRMSRSGSFRDGFEEVHG PEEKQSEIRKLRR++DASQE
 Sbjct: 3 SLDEKSRMSRSGSFRDGFEEVHGSSLSVSSSTSSIIYSTPEEKQSEIRKLRRDVDASQE 182
 Query: 1571 KVSALTTLTANAHLVAAFEQSLGNMTIRLQSLTMTAEQKDELNELRKTIEXXXXXXXXX 1630

LOCUS AA049124 337 bp mRNA EST 09-SEP-1996
DEFINITION mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
clone 479167 5'.
ACCESSION AA049124
NID gl528794
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mus.
REFERENCE 1 (bases 1 to 337)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-BHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-BHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:289911
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 292.
FEATURES
source Location/Qualifiers
1..337
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2]; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7T3 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."

/clone="479167"

/clone_lib="Soares mouse embryo NbME13.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

<1..>337

mRNA
 BASE COUNT 80 a 101 c 97 g 59 t
 ORIGIN

1 catcctctgt gggcaccgag gtcaccgaga cccctgctca ttcagtcccc cacactagac
 61 tggtccaagc caatgaagag gaggagccag agaagaagga ggtatcagaa ctgcgctctg
 121 aactatggga aaaagagatg aagctcacgg atatccggtt ggaggccctc aactctgccc
 181 accagctgga ccagcttcgg gagaccatgc acaatatgca gttggagggtg gacctgctga
 241 aagcagagaa tgaccggctg aaggttgccc cggggccctc ctcaggctgc actccagggc
 301 aggtccctgg gtcacgggt ctgtcgtccc ctcgacg

gb|AA049124|AA049124 mj46f04.r1 Soares mouse embryo NbME13.5 14.5 Mus
 musculus cDNA clone 479167 5'
 Length = 337

Plus Strand HSPs:

Score = 206 (92.9 bits), Expect = 3.9e-19, P = 3.9e-19

Identities = 42/60 (70%), Positives = 51/60 (85%), Frame = +3

Query: 1760 DSEAEVTVMLRNLRLDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLK 1819
 + E + V +LR+EL +KEMKLTDIRLEAL+SAHQLDQLRE M+ MQ E++ LKAENDRLK
 Sbjct: 84 EPEKKEVSELRLSELWEKEMKLTDIRLEALNSAHQLDQLRETMHNMQLVDLLKAENDRLK 263
 //

LOCUS AA185349 348 bp mRNA EST 07-JAN-1997
 DEFINITION mu51c03.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
 642916 5'.
 ACCESSION AA185349
 NID g1769059
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Mus.
 REFERENCE 1 (bases 1 to 348)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-BHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
 WashU-BHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:394908
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 336.

FEATURES Location/Qualifiers

source 1..348

/organism="Mus musculus"

/strain="C57BL/6J"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; [5' TGTTACCAATCTGAAGTGGGAGCGCGCGATACTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

/clone="642916"

/clone_lib="Soares mouse lymph node NbMLN"

/sex="male"

/dev_stage="4 weeks"

/lab_host="DH10B"

mRNA <1..>348

BASE COUNT 93 a 95 c 78 g 82 t

ORIGIN

```

1 attcggcact gaggggatga ataatccacc aaattagtgt gtacatagga gttgctgggc
61 cccccccac tcttatctgc tgtagctagc ctctccctaa gcctcgcac ttctctaaat
121 ctatctctgc gttcttacc cttgttctgg ccaatagaac tccggatcaa gaggcagaat
181 tcctcagata gcatctccag cctcaacagc atcaccagcc attccagcat cggcagcagc
241 aaagatgctg atgccaagaa gaaaaaagaag aagagttggg taagtaaagg cttggagata
301 ggctgtgct aggagtcact caccctgttg cagggaactg accccttt

```

//

gb|AA185349|AA185349 mu51c03.r1 Soares mouse lymph node NbMLN Mus
 musculus cDNA clone 642916 5'
 Length = 348

Plus Strand HSPs:

Score = 154 (69.4 bits), Expect = 4.4e-12, P = 4.4e-12
Identities = 27/42 (64%), Positives = 40/42 (95%), Frame = +1

Query: 1656 DLRIRRQHSDDSVSSINSATSESSVGSNIESDSKKKKRKNWL 1697
 +LRI+RQ+SSDS+SS+NS TSESS+GS+ ++D+KKKK+K+W+
Sbjct: 157 ELRIKRONSSDSISLNSITSESSIGSSKDADAKKKKKSWV 282

FIGURE 12a

"SIM output with parameters:
substitution scores in BLOSUM62
O = 12, E = 4"

Sequence 1: hu1, 1702 residues
Sequence 2: hu2, 2350 residues

List of local alignments with score >= 100.0

46.8% identity in 1726 residues overlap; Score: 2538.0; Gap frequency: 9.3%

```
hu1,      78 DPESQRRKRTVQNVLDRQNLLEETMSSLRGSQVTHSSLEMTCYDS--DDANPRSVSSLSNR
hu2,     639 DPEARLRLRTVKNIADLRQNLLEETMSSLRGTQVTHSTLETTFDTNVTTEMSGRSILSLTGR
          *** * *** * ***** * * * * * * * * * * * * * * * * * * * * * *
hu1,     136 SSPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMRSPSKLSHIS
hu2,     699 PTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLASRGSS
          ***** * * * * * * * * * * * * * * * * * * * * * *
hu1,     196 RLEL-VESLDSDEVDLKS-----GYMSDSLGMKMTEDDDITG-----
hu2,     759 VCHVDVSDKAGDEMDLEGISMDAPGYMSDGDVLSKNI-RTDDITSGYMTDGGGLGLYTRRL
          * * * * * * * * * * * * * * * * * * * * * *
hu1,     235 -----WDESSSISSGLSDASDNLSSSEEFNASSSLNSLP
hu2,     818 NRLPDGMVAVRETQRNTSLGLGDADSWDDSSSVSSGISDTIDNLTDDINTSSSISSYA
          * * * * * * * * * * * * * * * * * * * * * *
hu1,     268 STPTASRRNSTIVLRDSEKRSLAESGLSWFSEEEKAPKKLEYDSGLKMEPGTSKWRR
hu2,     878 NTPASSRKNLDV--QTDAEKHSQVERNSLWSGDDVKKSDGGS--DSG-IKMEPG-SKWRR
          ** * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,     328 ERPESCDDSSKGGELKKPISLGHGPGSLKKGKTPPVAVTSPITHTAQ--SALKVAGK---P
hu2,     932 NPSDVSDSDKSTSGRKNPVISQTGSWRRGMTAQVGITMPRTKPSAPAGALKTPGTGKTD
          * * * * * * * * * * * * * * * * * * * * * *
hu1,     383 EGKATDKGKLAVKNTGLQRSSSDAGRDRLSDAKKPPSGIARPSTSG--SFGYKKPP-PAT
hu2,     992 DAKVSEKGRLSPKASQVRSPSDAGRSSGDESKKPLPSSRTPTANANSFGFKKQSGSAA
          * * * * * * * * * * * * * * * * * * * * * *
hu1,     440 GTATVMQTG-----GSATLSKIQSSGIPVKPVNGRKTSLDVSNNAEPGFLAPGARSNIQ
hu2,    1052 GLAMITASGVTVTSRSATLGKIPKSSAL-VSRAGRKSSMDGAQNQDDGYLALSSRTNLQ
          * * * * * * * * * * * * * * * * * * * * * *
hu1,     495 YRSLPRPAKSSSMSVTGGRGGPRPVSSSIDPSLLSTKQGGGLTPSRLKEPTKVASGRTPA
hu2,    1111 YRSLPRPSKNSNRNGAGNRSS-----TSSID-SNISSKSAGLPVKLREPSKTALGSSSLPG
          ***** * * * * * * * * * * * * * * * * * * * * * *
hu1,     555 PVNQTDREKEKAKAKAVALDSNISLSKISGPESTPKNQASHPTATKLAELPPTPLRATA
hu2,    1166 LVNQTDKEKGISSDNESVASCNSVKVNPAAQPVSSPAQTSLOPGAKYPDVASPTLRLFLG
          ***** * * * * * * * * * * * * * * * * * * * * * *
```

hu1, 615 RSFVKPPSLANLDKV-NSNSLDLPSSSDTTHAS--KVPDLBATSSASGGPL-----P
hu2, 1226 GKPTKQVPIATAENMKNSVVISNPBATMTQQGNLDSPPSGSVLSSGSSSPLYSKNVDLNO
* * * * *
hu1, 664 SCFTPSPAPILNINSASFQGLELMSGFSVPKETRMYPKLSGLHRSMESLQMPMS---LP
hu2, 1286 SPLASSPSSAHSAPNSLWTGTAASSSSAVSKDGLGFQSVSSLHTSCESIDISLSSGGVP
* * * * *
hu1, 721 SAFPSSTPVPTPPAPPAAP-TEEETEELTWSGSPRAGQLDSNQ-----RD
hu2, 1346 SHNSSTGLIASSKDDSLTPFVRTNSVKTTLSSESPLSSPAASPKFCRSTLPRKQSDPHLD
* * * * *
hu1, 765 RNTLPKKGLRY----QLQSQEETKERRHSHETIGGLPESDDQSELSPPALPMSLSAKGQL
hu2, 1406 RNTLPKKGLRYTPTSQRLTQEDAKEWLRSHSAGGLQDTAANSFPSSGSSSVTSPSGTRFNF

hu1, 821 TNIVSPTAAT-----TPRITRSNSIPTHEAAFELYSGSQM-GSTLSLAERPCKMIRSGSF
hu2, 1466 SQLASPTTVTQMSLSNPTMLRTHSLSNADGQYDPTDSRFRNSSMSLDEKSRTMSRSGSF
* * * * *
hu1, 875 RDPTDDVHGSVLSLASSASSTYSSAEERMQSEQIRKLRRELESSQEKVATLTSQLSANAN
hu2, 1526 RDGFEEVHGSSSLVSSSTLSVYSTPEEKQSE-IRKLRRELDASQEKVSALTQTANAE
* * * * *
hu1, 935 LVAAFEQSLVNMTSRLRHLAETAEEKDTELLDLRETIDFLKKKNSEAOAVIQALNASET
hu2, 1585 LVAAFEQSLGNMTIRLQSLTMTAEQKDSSELNELRKTIELLKKQNAQAINGVINTPEL

hu1, 995 TPK-----ELRIKRONSSDSISSLSNITSHSSIGSSKDADAKKKKKSWVYELRSSF
hu2, 1645 NCKGNGTAQSADLRIRRHQSSDSVSSINSATSHSSVGSNIESDSKKKKRKNW---LRSSF
* * * * *
hu1, 1047 NKAFSIRKGPKSASSYSIDIEEATPDSSAPSSPKLQHGSTETASPSIKSSTLSSVGTDTV
hu2, 1702 KQAFGKKKSPKSASSHSDIE--TTDSSLPSSPKLPHNGSTGSTPLLRNHSNSL-----
* * * * *
hu1, 1107 EGPAPAPHTRLFHANESEEPEKKEVSELRLSELWEKEMKLTDIRLEALNSAHQLDQLRET
hu2, 1755 -----ISECMDSEAEVTVMQLRNELRDKEMKLTDIRLEALNSAHQLDQLREA
* * * * *
hu1, 1167 MBNMQLEVLDLKAENDRLKVAPGPSSGSTPGQVPGSSALS-SPRRSLGLALTHSFGPSLA
hu2, 1801 MNRMQSEIEKLKAENDRLK---SESQSGGCSRAPSQVSIASPRQSMGLS-QHSLNLTES
* * * * *
hu1, 1226 DTDLSPMDGISTCGPKEEVT--LRVVVRMPPQHIKGDLLKQOEFFLGCSKVSGKVDWKML
hu2, 1857 TSLDMLLDDTGECSARKEGGRHVKIVVSFQEMKWKEDSRPHLFLIGCIGVSGKTKWDVL
* * * * *
hu1, 1284 DEAVFQVFKDYISKMDPASTLGLSTESIBGYSISHVKRVLDAEPPPEMPPCRGVNN---I
hu2, 1917 DGVVRRLFKEYIIHVDPVSQLGLNSDVLGYSIGEIKRSNTSETPELLPCGYLVGENTTI
* * * * *

```
hu1,      1341 SVSLKGLKREKCVDSLVEFETLIPKPMQBYISLLKHRRLLVLSGPGSGTKTYLTNRLAEYL
hu2,      1977 SVTVKGLAENSLDSLVEFSLIPKPIQRYVSLLEHRRILSGPSGTGKTYLANRLSEYI
          ** *** * ***** ** * * * * * * * * * * * * * * * * * * * *
hu1,      1401 VERSGREVTEGIVSTFNMBQQSCDKLQLYLSNLANQIDRETGIGDVPLVILLDDLSEAGS
hu2,      2037 VLREGRELTGVIATFNVDHKSSKELRQYLSNLADQCNSENNAVDMPLVIILDNLHHVSS
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1461 ISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGLHLSFRMLTFSNNVEPANGFLVRYLR
hu2,      2097 LGEIFNGLLNCKYHKCPYIIGTMNQATSSSTPNLQLHHNFRWVLCANHTEPVKGFLGRFLR
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1521 RKLVESDSDINANKEELLRVLDWVFKLWYHLETFLEKHSSTDFLIGPCFFLSCPIGIEDF
hu2,      2157 RKLMEITEISGRVRNMELVKIIDWIPKVWHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGS
          *** * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1581 RTWFIDLWNNSIIPYLOEGAKDGKIVHGQKAAWEDPVEWVRDTLPWPSAQDQS--KLYH
hu2,      2217 RVWFTDLWNYSIIPYLLAEVREGQLYGRAPWEDPAKWVMDTYPWAASPOQHEWPPLLQ
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *
hu1,      1639 LPPPTVGPBSIASPPEDRTVKDSTPSSLDSDPLMAMLLKLQEAANY
hu2,      2277 LRPEDVGFDGYSMPREGSTSKQMPPSDAEGDPLMNMLMRLQEAANY
          * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

WARNING: 49 local alignments have not been reported because of score < 100.0

"SIM output with parameters:
substitution scores in BLOSUM62
O = 12, E = 4"

Sequence 1: Cel, 1583 residues
Sequence 2: hu2, 2350 residues

List of local alignments with score >= 54.0

32.8% identity in 504 residues overlap; Score: 490.0; Gap frequency: 6.9%

```
Cel,      1058 VIELKQELKERDSALYEVRLDNLDRAREVDVLRETVNKLTENKQLKKEVDKLTNGPATR
hu2,      1766 VMQLRNELRDKEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLEKENDRLKSESQGS
          * * * * * * * * * * * * * * * * * * * * * * * * * * * *
Cel,      1118 ASSRASIPVIYD-----DEHVDAAACSST-----SASQSSKRSSGCNSIKVTNVV
hu2,      1826 GCSRAPSQVSISASPRQSMGLSQHSLNLTESTSLDMLLDDTGECSARKEGGRHVKIVVSF
          *** * * * * * * * * * * * * * * * * * * * * * * * * * *
Cel,      1163 DIAGEISSIVNPDKEIIVGYLAMSTSQSCWKDIDVSIILGLFEVYLSRIDVEHQLGIDARD
```

hu2, 1886 QEEMKWKEDSRPHL-FLIGCIGVS-GKTKWDVLDGVVRRLFKEYIIHVDPVSQGLGNS-D
* * * * *
Cel, 1223 SILGYQIGELRRVIGDSTMTITSBPTDILTSSTTIRMFMHGAAQSRVDSLVDMLLPKQM
hu2, 1943 SVLGYSIGEIKRSNTSETPELLPCGY-LVGENTTISVTVKGLAENSLDSLVPESLIPKPI
* * * * *
Cel, 1283 ILQLVKSILTERRLVLGATGIGKSKLAKTLAAYVSIRTNQ--SEDSIVNISIPENKKEE
hu2, 2002 LQRYVSLLEHRRRIILSGPSGTGKTYLANRLSEYIVLREGRELTGVIATFNVDHKSKE
* * * * *
Cel, 1341 LLOVERRLEKIILRSKESCI-----VILDNIPKNRIAFVVSVFANV-PLQNEGPFVVCVCTV
hu2, 2062 LRQYLSNLADQCSENNAVDMPLVIILDNL--HBVSSLGEIFNGLLNCKYHKCPYIIGTM
* * * * *
Cel, 1395 NRY--QIPELQIHBNFKMSVMSNRLE---GFILRYLRRRAVEDEYRLTVQMPSELFKIID
hu2, 2120 NQATSSTPNLQLEHBNFRWVLCANBTEPVKGFGLGRFLRRKLMETEISGRVRN-MELVKIID
* * * * *
Cel, 1450 FFPIALQAVNNFIEKTSVDVTGPRACLNCLPTVDGSRWFIRLWNNFIPYLERVARD
hu2, 2179 WIPKVVHHLNRFLEAHSSSDVTIGPRLFLSCPIDVDGSRVWFIDLWNYSIIPYLLEAVRE
* * * * *
Cel, 1510 GKRTFGRCTSFEDPTDIVSEKWPW
hu2, 2239 GLQLYGRRAFWEDPAKWVMDTYPW
* * * * *

35.5% identity in 112 residues overlap; Score: 165.0; Gap frequency: 1.8%

Cel, 11 IYTDWANRHLKSGSLKSIIRDISNDFRDYRLVSQLINVIPINEFSPAFTKRLAKITSNL
hu2, 11 IYTDWANHYLTKSGHKRLIKDLQDDVTDGVLLAQIIQVVA--NEKIEDINGCPKNRSQMI

Cel, 71 DGLETCLDYLKNLGLDCSKLTKTDIDSGNLGAVLQLLFLLSTYKQKLRQLKK
hu2, 69 ENIDACLNFLAAKGINIQGLSAEEIRNGNLKAILGLFFSLSRKQKQQQQPQK
* * * * *

24.8% identity in 163 residues overlap; Score: 80.0; Gap frequency: 3.7%

Cel, 877 GSQSLASTT--AYGSLNEKYEHAIKRDMDARDLECYKNTVDLSLTKKQENYGALFDLFEQKL
hu2, 1534 GSSLSLVSTLSVYSTPEEKQSEIRKLRRELDASQEKVSALTTQLTANAHLVAAFEQSL
* * * * *
Cel, 935 RKLTOHIDRSNLKPEEAIRFRQDIAHLRDISNHLASNSAHANEGAGELLRQPSLESVASH
hu2, 1594 GNMTIRLQSLTMTAEQK---DSELNELRKTIELKKQNAQAQAINGVINTPELNCKGNG
* * * * *
Cel, 995 RSSMSSSSKSSKQEKISLSSFGK-NKKSWIRSSLSKFTKKKNK
hu2, 1651 TAQSADLRIRRHSSDSVSSINSATSHSSVGSNIESDSKKKKR
* * * * *

58.6% identity in 31 residues overlap; Score: 74.0; Gap frequency: 6.5%

F-----
Cel, 653 GYPDNFEDSSSLSSGISDNNELDDISTDDL
hu2, 840 GDADSWDDSSSVSSGISDT--IDNLSTDDIN
* * * * * * * * * *

42.9% identity in 60 residues overlap; Score: 64.0; Gap frequency: 6.7%

Cel, 984 RQPSLESVASHRSMSSSSKSSKQEKISLSSFGKNKSWIRSSLK-FTRKKKNKYDEAH
hu2, 1661 RQBSSDSVSSINSATSHSSVGS---NIESDSKKKKRKNWLRSSFKQAFGKKKSPKSASSH
* * * * * * * * * * * * * * * * * * * *

22.0% identity in 91 residues overlap; Score: 56.0; Gap frequency: 0.0%

Cel, 140 SKLPSPRVATSATASATNPNSNFPQMSTSRLOTPQSRISKIDSSKIGIKPKTSGLKPPSS
hu2, 177 SRLSGPTARVSAAGSEAKTRGGSTANNRRSQSFNNYDKSKPVTSPPPPPSSHEKEPLAS
* * * * * * * * * * * * * * * * * * *

Cel, 200 STTSSNNTNSFRPSSRSSGNNNVGSTISTSA
hu2, 237 SASSHPGMSDNAPASLESGSSSTPTNCSTSS
* * * * * * * * * *

WARNING: 44 local alignments have not been reported because of score < 54.0

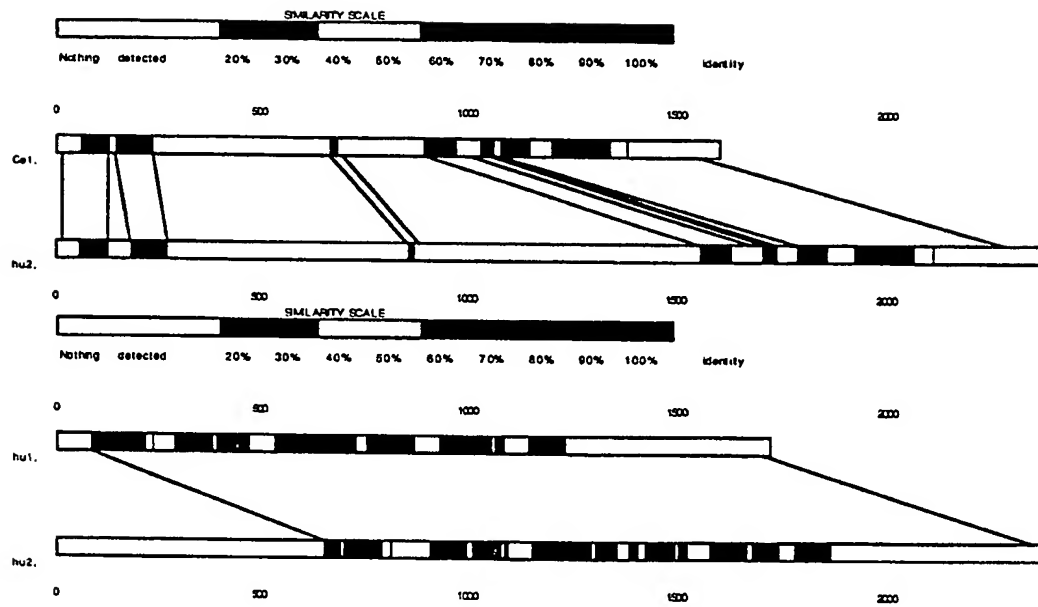


Figure 12b

Tuesday, 18 November 1997 10:09

fig 13 pCB201 (1 > 5082) Site and Sequence

Enzymes : 100 of 148 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page

fig 13 6/24/97

GACGGATCGGGAGATCTCCCGATCCCCATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGT
CTGCTAGCCCTCTAGAGGGCTAGGGGATACCAGCTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTGGTTCATAGACGAGGGACGAACACACAA
T D R E I S R S P M V D S Q Y N L L . C R I V K P V S A P C L C V 100

GGAGGTCGCTGAGTAGTCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTGGG
CCTCCAGCGACTCATCACGCGCTCGTTTTAAATTCGATGTTGTTCCGTTCCGAACCTGGCTGTTAAGCTACTTCTTAGACGAATCCCAATCCGCAAAACGG
G G R . V V R E O N L S Y N K A R L D R O L H E E S A . G . A F C 200

CTGCTTCGCGATGTACGGGCGAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
GACGAAGCGCTACATGCCGGTCTATATGCGCAACTGTAACATAAATGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATAT
A A S R C T G O I Y A L T L I I D . L L I V I N Y G V I S S . P I Y 300

TGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
ACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTGTCTGGGGGGGGTAACGTCAGTTATTACTGCATACAAGGGTATCA
G V P R Y I T Y G K V P A V L T A Q R P P P I D V N N D V C S H S 400

AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTACGCCC
TTGCGGTTATCCCTGAAAGGTAAGTGCAGTTACCCACCTGATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTTATGCGGG
N A N R D F P L T S M G G L F T V N C P L G S T S S V S Y A K Y A 500

CCTATTGACGTCAATGACGTAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTCTTACTTGGCAGTACATCTACGTATTAGTCA
GGATAACTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTGATGACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAG
P Y . R O . R . M A R L A L C P V H D L M G L S Y L A V H L R I S H 600

TCGCTATTACCATGGTGATGGGTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTGACTCAGGGGATTTCGAAGTCTCCACCCCATTGACGTCAA
AGCGATAATGGTACCCTACGCCAAAACCGTTCATGTAGTTACCGGCACCTATCGCCAAACTGAGTGCCCTAAAGGTTACAGAGGTGGGGTAAGTGCAGT
P Y Y H G D A V L A V H O W A V I A V . L T G I S K S P P H . R Q 700

TGGAGTTTGTGTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAAC TCCGCCCATTTGACGCAAAATGGGCGGTAGCGGTGACGTTGGAG
ACCTCAAAACAAAACCGTGTGTTTGTGCTGAAAGGTTTACAGCATGTTGAGGCGGGGTAACGCGTTTACCGGTCATCCGCACATGCCACCTC
W E F V L A P K S T G L S K M S . Q L R P I D A N G R . A C T V G 800

GTCATATTAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCAAGCTGGCTAGC
CAGATATATTGCTCGAGAGACCGATTGATCTCTGGGTGACGAATGACCGAATAGCTTTAATTATGCTGAGTGATATCCCTCTGGGTTCGACCGATCG
900

G L Y K G S S L A N . R T H C L L A Y R N . Y D S L . G D P S V L A

GTTTAACTTAAGCTTACCATGGGGGTTCTCATCATCATCATCATGGTATGGCTAGCATGACTGGTGGACAGCAAAATGGGTCGGGATCTGTACGAG
CAAAATTGAATTCGAATGGTACCCCCAAGAGTAGTAGTAGTAGTACCATACCGATCGTACTGACCACTGTGCTTTACCCAGCCCTAGACATGCTG
1000

F K L K L T M G G S H H H H H G M A S M T G G Q Q M G R D L Y D

T7 promoter priming site

ProBond binding domain

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GATGACGATAAGGTACCTAGGATCCATATGCCCTCTTGGCGTCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGCTGAAGGAGAAATGCGTCGACAT
CTACTGCTATTCCATGGATCCTAGGTATACGGAGGAACGGCAGCTCCACATTTATGTATAGTCAGAGGGAGTTTCCAGACTTCTCTTTACGCAGCTG
pCB201 insert = U4
U4 ORF
D D D K V P R I H M P P C R R G V N N I S V S L K G L K E K C Y D
GCCGTGGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAASCTCTGCTGAAGCACCGGCGCCTCGTCTCTCGGGCCCCAGCGGAC
CGGACCACAAGCTCTGCGACTAGGGGTTCCGGCTACTACGTCGTGATGTATTCGGAGGACGACTTCGTGGCCGCGGAGCAGGAGAGCCCCGGGTCGCCGTS
pCB201 insert = U4
U4 ORF
S L V F E T L I P K P M M Q H Y I S L L L K H R R L V L S G P S G T
GGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTACAGAGGGCATCGTCAGCACCTTCAACATGCACCA
CCCGTTCGGATGGACTGGTTAGCGAACC GGCTCATGGACCACCTCGCGAGACCGGCACTCCAGTGTCTCCCGTAGCAGTCGTGGAAGTTGTACGTGGT
pCB201 insert = U4
U4 ORF
G K T Y L T N R L A E Y L V E R S G R E V T E G I V S T F N M H Q
CASTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACCAAGATACACCGGGAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATG
GTGAGAACGTTCTAGACGTTGACATAGAAAGGTTGGATCGGTTGGTCTATCTGGCCCTTTGTCTTAACCCCTACACGGGGACCACTAAGATAACCTAC
pCB201 insert = U4
U4 ORF
U S C K D L Q L Y L S N L A N O I D R E T G I S D V P L V I L L D
ACCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCATATTATAGGTACCAACCAATCAGCTG
TGGATCACTTCGTCCGAGGTAGTCACTCAACCAAGTTACCCCGGGAGTGGACGTTTATAGTATTTACAGGATATAATATCCATGGTGGTTAGTGGAC
pCB201 insert = U4
U4 ORF
D S E A G S I S E L V A G A L T C K Y H K C P Y I I G T T N Q P V
AAAAATGACACCCAACCATGGCTTGCACTTGAGCTTCAGSATGTTGACCTTCTCCAACAACGTTGGAGCCAGCCAAATGGCTTCCTGGTTGCTTACCTGAG
TTTTTACTGTGGGTTGGTACCGAACGTGAACCTGSAAGTCTACAAC TGAAGAGGTTGTTGCACCTCGGTTCGGTTACCGAAGGACCAAGCAATGGACTCC
pCB201 insert = U4
U4 ORF
K M T P N H G L H L S F R M L T F S N N V E P A N G F L V R Y L F

BASE COUNT	173 a	168 c	141 g	128 t		
ORIGIN						
1	gggccctcta	gggtgcctgc	tgccaggaagc	acagcatagg	tccagggagc	ctctaattta
61	aataggagaa	gtcagagctt	taacagcatt	gacaaaaaca	agcctccaaa	ttatgcnaat
121	ggaaacgaaa	aagattcctc	caaaggacct	caatcgtctt	caggtgtaaa	tggtaacgtg
181	cagcctccca	gtactgctgg	gcagcctcct	gcctctgccca	tcccttctcc	aagtgccagc
241	aagccctggc	gcacgaagtc	catgaatgtc	aaacacagtg	ccacctccac	catgttgact
301	gtaaagcagt	caagtacagc	cacctcccc	acaccatctt	cagacagact	gaaggcaacc
361	tgtctcagaa	ggggtcaaaa	ctgctccctc	aggacagaaa	tccatgcttg	agaaattcaa
421	gctagtcaat	gcccggactg	ctttacgccc	cccgagcct	cccagttcag	gacctagtga
481	tggtgggaag	gatgatgatg	cottttctga	atctggtgaa	atggaagggt	ttaacagtgg
541	tctgaatagt	ggtggctcaa	caaatagcag	tcccaaagtg	tcacctaaagt	tggtccctcc
601	aaaagctgga					

LOCUS AA495042 418 bp mRNA EST 27-JUN-1997
 DEFINITION fa05f06.s1 Zebrafish ICRFzfls Danio rerio cDNA clone 5D16 3'.
 ACCESSION AA495042
 NID g2225470
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
 Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Ostariophysi; Cypriniformes; Cyprinoidea; Cyprinidae; Rasborinae;
 Danio.
 REFERENCE 1 (bases 1 to 418)
 AUTHORS Clark,M., Lehrach,H., Johnson,S., Marra,M., Eddy,S., Billier,L.,
 Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T.,
 Lacy,M., Le,N., Lennon,G., Martin,J., Moore,B., Schellenberg,K.,
 Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R.
 and Wilson,R.
 TITLE WashU Zebrafish EST Project
 JOURNAL Unpublished (1997)
 COMMENT
 Contact: Steve Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Steve Johnson lab internal ID - P2_60 NOTE - For this library, the
 CLONE id field represents a position identifier on the original
 cDNA library preparation plate. cDNA Library Preparation: Matthew
 Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by:
 Washington University Genome Sequencing Center Clone distribution:
 Genome Systems, St. Louis, and Max Planck Institut fuer Molekulare
 Genetik, Berlin Tel +49 30 84 13 1235
 Seq primer: -40ml3 ET from Amersham
 High quality sequence stop: 416.
 FEATURES
 source Location/Qualifiers
 1..418
 /organism="Danio rerio"
 /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
 strand cDNA was primed with a Not I - oligo(dT)15 primer
 [5'pGACTAGTTCTAGATCGGAGCGGCCGCTTTTTTTTTTTT3'], on
 mRNA from pooled 26 somite zebrafish embryos;
 double-stranded cDNA was ligated to Sal I adaptors (BRL),
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max
 Planck Institut fuer Molekulare Genetik, Berlin) and was
 not biochemically normalised. 70,000 clones from this
 library were arrayed on high density filters and
 subsequently screened by oligonucleotide hybridization
 fingerprinting to identify unique or minimally redundant
 clones for more intensive analysis."

205/270

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

AGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAGGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCT
TCC TTCGACCATCTCAGTCTGTGCTGTAGTTACGGTTGTTCTTCGACGAAGCCACGAGCTGACCCATGGGTTCGACACCATAGTAGAGGTGTGGG

pCB201 insert = U4

U4 ORF

R K L V E S D S D I N A N K E E L L R V L D V V P K L V Y H L H T

TCCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTCTGTGCTGTCCATTGGCATTGAGGACTTCGGGACCTGCTTCATTGACCT
AGGAACTCTTCGTGCTGAGTCTGAAGGAGTAGCCGGGAACGAAGAAAGACAGCACAGGGTAACCGTAACCTCTGAAGGCCGAGGACCAAGTAAC TGGG

pCB201 insert = U4

U4 ORF

F L E K H S T S D F L I G P C F F L S C P I G I E D F R T V F I D L

GTGGAACAACTCTATCATTCCCTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAA TGGT
CACCTTGTGAGATAGTAAGGGATAGATGTCTTCTCGGTTCTTACCTATTTCAGGTACCTGTCTTTCGACGAACCTCTCGGGTCACCTTACCCAG

pCB201 insert = U4

U4 ORF

V N N S I I P Y L Q E G A K D G I K V H G Q K A A W E D P V E V V

CGGGACACACTTCCTTGGCCATCAGCCCAACAAGACCAATCAAGCTGTACCACTGCCCCACCCACCGTGGGCCCTCACAGCATTGCTCAGCTCCG
GCCCTGTGTGAAGGGACCGGTAGTCGGGTTGTTCTGGTTAGTTTCGACATGGTGGACGGGGTGGGTGGCACCCGGGAGTGTCTGAACGGAGTGGAGGGG

pCB201 insert = U4

U4 ORF

R D T L P V P S A Q Q D Q S K L Y H L P P P T V G P H S I A S P P

AAGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGAAGCTGCCAACTACAT TGAAT
TCCTATCTTGTGAGTTTCTGCTGCTGGGGTTCAAGAGACCTGAGTCTAGGAGACTACCGGTACGACGACTTGAAGTTCTTCGACGTTGATGTAAC TCA

pCB201 insert = U4

U4 ORF

E D R T V K D S T P S S L D S D P L M A M L L K L Q E A A H Y I E S

TCCAGATCGAGAAACCATCTCGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCAC TGTACCCCGGACAGCAGAACGCTGGCATCAGCTA
AGGTCTAGCTCTTTGGTAGGACCTGGGGTTGGAAGTCCGTTGTGAATTCCTAAGCCGTTAGTGACAGTGGGGGCCGTGCTGCTTTCGACCGTAGTGGG

pCB201 insert = U4

U4 ORF

P D F E T I L D P N L O A T L G F G N H C H P R T A E R N H Q L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

TCTTAGCTCCTCCTCTCCCCTCCTCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGAGATGAAGAGGAGGGGACAGGTT
AGAATCGAGGAGGAGAGGGGAGAGGAGAAAGTCTCGTGACCGAGAGGTGCGGGTCTCCTCTTGCTCCTCCTCCTCTACTTTCTCCTCCTGTCCAA 2300

pCB201 insert = U4

S . L L L S P L L F Q S T G S P A P G G E Q E G G G D E R G G T G

CTTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACCTTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACT
GAACCACGACATGGAACTCTTGAAGGATCCTTCTTACCACCCACCGCAAAACCTTGAACACGGGGGATTTGTGTAATGACCGGAGGAGATTACTGA 2400

pCB201 insert = U4

S V C C T F E N F L G R N G G V A F G N L C P L N T F T G L L . . L

TTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCTTGTTCAATTACAACTCCTGGGCTTTCTGGGGAGGGGTTTCAGAAAACATCAAAACACTGC
AACCCCTTTTCTACTAAGACCCAGAAAGGGAAGTGAAGAACAAAGTTAATGTTTGGAGACCGAAAGACCCCTCCCAAGTCTTTTGTAGTTTGTGACG 2500

pCB201 insert = U4

V G K D D S G S F P . L L V S I T N S V A F V G G V Q K T S K H C

AGCAGTTCTTAATGATTCTACAAGCAACCTGAGAGAGACAGTCTTGTGAGGGAGATCTGGGGAGGCGAGGAAGCTCTCAGATTTTCTCACAGACCC
TCGTCAAGGATTTACTAAGAGTGTTCTGTTGGGACTCTCTCTGTGACAACTCCCTCTAGACCCCTCCGTCTTCGAGGAGTCTAAAAGAGTGCTGGG 2600

pCB201 insert = U4

S S S . M I L T S N P E R D S L V R E I V G R Q E A P O I F S Q T

TTCCCAATTCATCACCACGCAACAC TCGTCCGGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCGTTTAAACCCGCTG
AAGGTTAAGGTAGTGGTGACGGTTGTGAGCAGCCTTAAGACGCTATAGSTCGTGTACCCGCCGCGAGCTCAGATCTCCCGGGCAAAATTGGGGCAG 2700

pCB201 insert = U4

L P N S I T T A N T R P E F C R Y P A Q V R P L E S R G P V . T R .

ATCAGCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCTTCCCCGTCCTTCTTGACCCGGAAGGTGCCACTCCCACTGTCTTTCC
TAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAACGGGGAGGGGACGGAAGGAAC TGGGACCTTCCACGGTGAGGGTGACAGGAAAGG 2800

S A S T V P S S C Q P S V V C P S P V P S L T L E G A T P T V L S

TAATAAATGAGGAAATGCATCGCATTGTCTGAGTAGGTGTCTATTCTATCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATTATTTTACTCCTTTAACGTAGCGTAACAGACTCATCCACAGTAAGATAAGACCCCCACCCACCCCGTCTGTCTGTTCCCCCTCCTAACCTTCTGT 2900

. . N E E I A S H C L S R C H S I L G G G V G Q D S K G E D V E D

ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGG
TATCGTCCGTACGACCCCTACGCCACCCGAGATACCGAAGACTCCGCTTTCTTGGTCGACCCGAGATCCCCCATAGGGGTGCGCGGGACATCGCGGGS 3000

N S R H A G D A V G S M A S E A E R T S V G S R G Y P H A P C S G A

ATTAAGCGCGGGGGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCGCTCTTTCGCTTTCTTCCCTTCTTCTCGG
TAATTCGCGCGCCACACCAATGCGGTGCGCTGCGGATGTGAACGCTCGGGGATCGCGGGCAGGAAAGCGAAAGGAAGGAAAGAGAGCGS 3100

L S A A G V V V T R S V T A T L A S A L A P A P F A F F P S F L A

ACGTTCCGCGGCTTTCCCCGTCAAGCTCTAAATCGGGCATCCCTTTAGGGTTCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACCTTGATTAGS
TGCAAGCGCGCGAAAGGGCAGTTCGAGATTTAGCCCGTAGGGAATCCCAAGGCTAAATCAGAAATGCCGTGGAGCTGGGGT TTTTGAAC TAATCC 3200

T F A G F P R Q A L N R G I P L G F R F S A L R H L D P K K L D .

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

GTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCTCAAACTGG
CCTACCAAGTGCATCACCCGGTAGCGGACTATCTGCCAAAAAGCGGAACTGCAACCTCAGGTGCAAGAAATTATCACCTGAGAACAGGTTTGACC
G D G S R S G P S P . . T V F R P L T L E S T F F N S G L L F Q T G
AACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAAGGGATTTGGGGATTTCGGCCTATTGGTTAAAAATGAGCTGATTTAACAAAAATTT
TTGTTGTGAGTTGGGATAGAGCCAGATAAGAAAC TAAATATTCCTTAAACCCCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTTGTTTTAA
T T L N P I S V Y S F D L . G I L G I S A Y V L K N E L I . Q K F
AACGCGAATTAATTCGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGT
TTSCGCTTAATTAAGACACCTTACACACAGTCAATCCACACCTTTTCAGGGGTCCGAGGGGTCCGTCCTTCATACGTTTCGTACGTAGAGTTAATCA
N A N . F C G M C V S . G V E S P Q A P Q A G R S M Q S M H L N .
CAGCAACCAGGTGTGGAAAGTCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTACGCAACCATAGTCCCGCCCTAACTCC
GTCGTTGGTCCACACCTTTTCAGGGGTCCGAGGGGTTCGTCCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCCGGGATTGAGG
S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P L T P
GCCATCCCGCCCTAACTCCGCCAGTTCCGCCCATTCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCT
CGGGTAGGGCGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCAGCTGATTAAAAAAATAAATACGTCCTCCGGCTCCGGCGGAGACGGAGA
P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A S A S
GAGCTATTCAGAAGTAGTGAAGAGGCTTTTTTGGAGGCCATAGGCTTTTGCAAAAAGTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGA
CTCGATAAGGTCTTCATCACTCTCCGAAAAAACCCTCGGATCCGAAAAAGTTTTTCGAGGGCCCTCGAACATATAGGTAAAGCCCTAGACTAGTTCTCT
E L F Q K . . G G F F G G L G F C K K L P G A C I S I F G S D Q E
CAGGATGAGGATGTTTCGATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACA
GTCTACTCTAGCAAGCGTACTAATTGTTCTACCTAACGTGCGTCCAAGAGGCCGGCAACCCACCTCTCCGATAAGCCGATCTGACCCGTTGTG
T G . G S F R M I E Q D G L H A G S P A A V V E R L F G Y D W A Q Q
GACAATCGCTGCTCTGATGCCGCCGTGTTCCGGCTGTGAGCGCAGGGGCCGCCGGTCTTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAAGT
CTTTAGCCGAGGAGACTACGGCGGCACAAGGCCGACAGTCGCTCCCGCGGGCCAAGAAAAACAGTTCTGGCTGGACAGGCCACGGGACTTACTTGAC
T I G C S D A A V F R L S A Q G R P V L F V K T D L S G A L N E L
CAGGACGAGGACGGCGGCATCTGTGGCTGGCCACGACGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAAGGAGCTGGCTGCTAT
GTCTGCTCCGTCGCGCCGATAGCACCGACCGGTGCTGCCCGCAAGGAACGCGTCGACACGAGCTGCAACAGTGACTTCGCCCTTCCTGACCGACGAT
G D E A A R L S V L A T T G V P C A A V L D V V T E A G R D V L L
TGGGCGAAGTGCCGGGCGAGGATCTCTGTATCTACCTTTGCTCTCGCGAGAAATATCCATCATGGCTGATGCAATGCGGCGGCCTGCATACGCTTGA
ACCGGCTTCACGGCCCCGTCTAGAGGACAGTAGAGTGAACGAGGACGGCTCTTTCAAGGTAGTACCGACTACGTTACGCCGCCGACGTATGCGAAG
L S E V P G Q D L L S S H L A P A E K V S I M A D A M R R L H T L D
TCGGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGACGACTCTCGGATGGAAGCCGGTCTTGTGATCAGGATGATCTGGACGAA
AGGCGGATGGACGGTAAGCTGGTGGTTGCTTTGTAGCGTAGCTCGCTCGTGCATGAGGCTACCTTCGGCCAGAACAGCTAGTCTCTACTAGACCTGCTT
P A T C P F D H O A K H R I E R A R T R M E A G L V D Q D D L D E
GAGCATCAGGGGCTCGCGCCAGCCGAACGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCTGTGACCTATGGCGATCTCTGCTTGG
CTGCTAGTCCCCGAGCGCGGTGGCTTGACAAGCGGTCCGAGTTCGCGCGTACGGGCTGCCGCTCTAGAGCAGCACTGGGTACCGCTACCGACGAAGC
E H G G L A P A E L F A R L K A R M P D G E D L V V T H G D A C L

Tuesday, 18 November 1997 10:09
fig 13 pCB201 (1 > 5082) Site and Sequence

Page

CGAATATCATGGTGGAAAAATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCCTATCAGGACATAGCGTTGGCTACCCGTGA
GCTTATAGTACCACCTTTTACCGGCGAAAAGACCTAAGTAGCTGACACCGGCCGACCCACACCGCTGGCGATAGTCTGTATCGCAACCGATGGGCACCT 4500
P N I M V E N G R F S G F I D C G R L G V A D R Y Q D I A L A T R D
TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTT
ATAACGACTTCTCGAACCGCCGCTTACCCGAC TGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAA 4600
I A E E L G G E V A D R F L V L Y G I A A P D S Q R I A F Y R L L
GACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCAGCGCCCAACCTGCCATCACGAGATTTGATTCCACCGCCGCTTCTATGA
CTGCTCAAGAAGACTCGCCCTGAGACCCCAAGCTTTACTGGCTGGTTCCGTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACT 4700
D E F F . A G L V G S K . P T K R R P T C H H E I S I P P P S M
AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATT
TTCCAACCCGAAGCCTTAGCAAAAGGCCCTGCGGCCGACCTACTAGGAGGTCGCGCCCTAGAGTACGACCTCAAGAAGCGGGTGGGGTTGAACAAATAA 4800
K G V A S E S F S G T P A G . S S S A G I S C V S S S P T P T C L L
GCAGCTTATAATGGTTACAAATAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACCTGCATTCTAGTTGTGGTTTGCCAAAAGTCATCA
CGTCGAATATTACCAATGTTTATTTTCGTTATCGTAGTGTTTAAAGTGTTTATTTTCGTAAAAAAGTGACGTAAGATCAACACCAACAGGTTTGAGTAGT 4900
Q L I M V T N K A I A S Q I S Q I K H F F H C I L V V V C P N S S
ATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATGTTTATCCGCTCACAA
TACATAGAATAGTACAGACATATGGCAGCTGGAGATCGATCTCGAACCGCATTAGTACAGTATCGACAAAGGACACACTTTAACAATAGCGAGTGTTA 5000
M Y L I M S V Y R R P L A R A V R N H G H S C F L C E I V I R S Q
TCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG 5082
AGGTGTGTTGTATGCTCGGCCCTTCGTATTTACATTTTCGGACCCACGGATTACTCACTCGATTGAGTGTAATTAACGCAAC
F H T T Y E P E A . S V K P G V P N E . A N S H . L R W

209/270

FIG. 14.

v360 v370 v380 v390 v400 v410 v420 v430 v440
SSVGTDVTEGPAHPAPHTRLPFANKEKEPEKKEKVSILRSELWEKEMKLTDIRLEALNSAQDQLRETHENMQLEV DLLKAENDRLKVAP
SSVGT:VTE.PAH:..PSTRLPF:ANKEKEPEKKEKVSILRSELWEKEMKLTDIRLEALNSAQDQLRETHENMQLEV DLLKAENDRLKVAP
SSVGTETVTETPAHSVPSTRLPFQANKEKEPEKKEKVSILRSELWEKEMKLTDIRLEALNSAQDQLRETHENMQLEV DLLKAENDRLKVAP
^10 ^20 ^30 ^40 ^50 ^60 ^70 ^80 ^90

v450 v460 v470 v480 v490 v500 v510 v520 v530
GPFSSGSTPGQVPGSSALSSPFRSLGLALHESFGPSLADTDLSPMDGISTCGPKKEVTLRVVVRMPQBHIKGD LKQOEFFLGCSKVSGKV
GPSSG.TPGQVPGSSALSSPFRSLGLAL:H:F:PSL:DTDLSPMDGISTCG:KEVTLRVVVRMPQBHIKGD LKQOEFFLGCSKVSGKV
GPFSSGCTPGQVPGSSALSSPFRSLGLALSFPSPSLTDTDLSPMDGISTCGSKKEVTLRVVVRMPQBHIKGD LKQOEFFLGCSKVSGKV
^100 ^110 ^120 ^130 ^140 ^150 ^160 ^170 ^180

v540 v550 v560 v570 v580 v590 v600 v610 v620
DWRHLD EAVFQVFKDYISKMDPASTLGLSTESIEGYSISEVVRVLDARPEMPPCRRGVNNISVSLKGLKEKCVDSL VFTLIPKPMQH
DWRHLD EAVFQVFKDYISKMDPASTLGLSTESIEGYS:SEVVRVLDARPEMPPCRRGVNNISV:LKGLKEKCVDSL VFTLIPKPMQH
DWRHLD EAVFQVFKDYISKMDPASTLGLSTESIEGYSLSHVVRVLDARPEMPPCRRGVNNISVALKGLKEKCVDSL VFTLIPKPMQH
^190 ^200 ^210 ^220 ^230 ^240 ^250 ^260 ^270

v630 v640 v650 v660 v670 v680 v690 v700 v710
YISLLKHRRRLVLSGPGSGTGKTYLTNRLAAYLVERSGREVTGIVSTFNMEQQSCDKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
YISLLKHRRRLVLSGPGSGTGKTYLTNRLAAYLVERSGREVT:GIVSTFNMEQQSCDKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
YISLLKHRRRLVLSGPGSGTGKTYLTNRLAAYLVERSGREVTGIVSTFNMEQQSCDKDLQLYLSNLANQIDRETGIGDVPLVILLDDLSEA
^280 ^290 ^300 ^310 ^320 ^330 ^340 ^350 ^360

v720 v730 v740 v750 v760 v770 v780 v790 v800
GSISELVNGALTCYKCPYIIGTTNQPVKHTPNHGLELSFRMLTFSNNVEPANGFLVRYLRRLKLVESDSDINANKELLRLVDWVPKLW
GSISELVNGALTCYKCPYIIGTTNQPVKHTPNHGLELSFRMLTFSNNVEPANGFLVRYLRRLKLVESDSD:NA NKELLRLVDWVPKLW
GSISELVNGALTCYKCPYIIGTTNQPVKHTPNHGLELSFRMLTFSNNVEPANGFLVRYLRRLKLVESDSDVNA NKELLRLVDWVPKLW
^370 ^380 ^390 ^400 ^410 ^420 ^430 ^440 ^450

v810 v820 v830 v840 v850 v860 v870 v880 v890
YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTWIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQDQSKLYH
YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTWIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQDQSKLYH
YHLSTFLKXSTSDFLIGPCFFLSCPIGIEDFRTWIDLWNNSIIPYLQEGAKDGIKVHGQKAAWEDPVEWVRDTLPWPSAQDQSKLYH
^460 ^470 ^480 ^490 ^500 ^510 ^520 ^530 ^540

v900 v910 v920 v930 v940 v950 v960
LPPFTVGFPHSIASPPEDRTVKDSTPSSLSDSLPHAMLLKLQEAANYIESPDRETILDPNLQATL
LPPF:VGFPHS.ASPPEDRTVKDSTP:SLDSLPHAMLLKLQEAANYIESPDRETILDPNLQATL
LPPFVGFPHSTASPPEDRTVKDSTPNSLSDSLPHAMLLKLQEAANYIESPDRETILDPNLQATL
^550 ^560 ^570 ^580 ^590 ^600

210/270

FIG. 15.

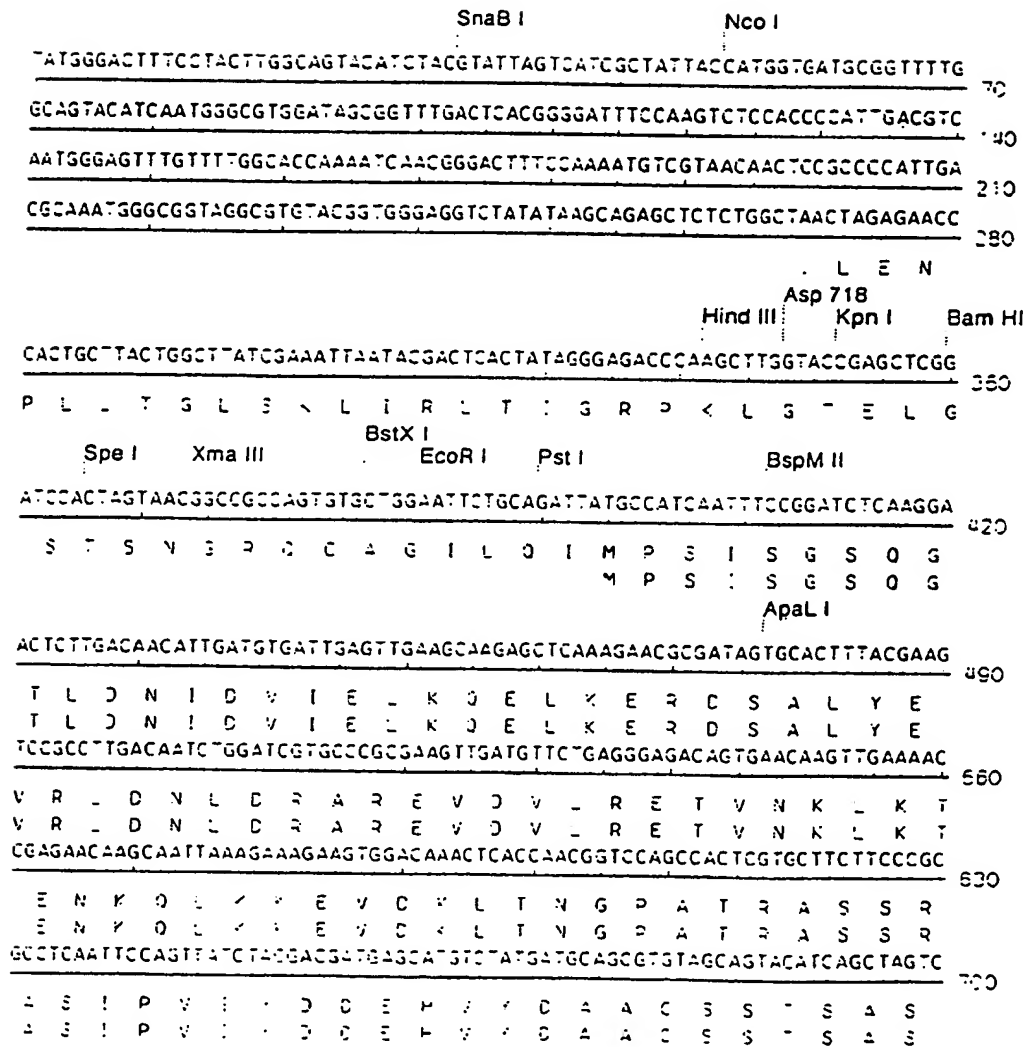


FIG. 15 CONTINUED.

Asu II

AATCTTCGAAACGATCCTCTGGCTGCAACTCAATCAAGTTACTGTAAACGTGGACATCGCTGGAGAAAT 770
C S S K R S S G C N S I K V T V V V D I A G E :
C S S K R S S G C N S I K V T V V V D I A G E :

Pvu I

Hpa I

EcoR V

CAGTTCGATCGTTAACC CGSACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCAGTCAGTCATGC 840
S S I V N P D K E I I V G Y L A M S T S Q S C
S S I V N P D K E I I V G Y L A M S T S Q S C
TGGAAAGACATTGATGTTCTATTCTAGGACTATTGAAGTCTACCTATCCAGAATTGATGTGGAGCATC 910

V K D : D V S I L G L F E V Y L S R I C V E H
V K D : D V S I L G L F E V Y L S R I C V E H

Cla I

Mlu I

AACTTGGAAATCGATGCTGCGTATTCTATCCTTGGCTATCAAATTGGTGAACCTCGACCGCTCATTGGAGA 980
Q L G : D A R D S I L G Y Q I G E L R R V I G D
Q L G : D A R D S I L G Y Q I G E L R R V I G D
CTCCACAACCATGATAACCGCCATCCAACGTGACATTCTTACTTCCCAACTACAATCCGAATGTTCTATG 1050

S T T M : T S H P T D : L T S S T T I R M F M
S T T M : T S H P T D : L T S S T T I R M F M
CACGCTGCCGCACAGAGTCCGCTAGACAGTCTGGTCTTGATATGCTCTTCCAAAGCAAATGATTCTCC 1120

H G A A Q S R V D S L V L D M L L P K C M I L
H G A A Q S R V D S L V L D M L L P K C M I L
AACTCGTCAAGTCAATTTTGACAGAGAGACGTC TGGTGTAGCTGGAGCAACTGGAAATGGAAAGAGCAA 1190

Q L V K S : L T E R R L V L A G A T G I G K S K
Q L V K S : L T E R R L V L A G A T G I G K S K

Asu II

ACTGGCGAAGACCTTGGCTGCTTATGTATCTATTGGAACAAATCAATCCGAAGATAGTATGTGAATATC 1260
L A K T L A A V S I P T N Q S E C S : V N I
L A K T L A A V S I R T N Q S E C S : V N I

Bsm I

Bgl II

AGCAATTCCTGAAAACAATAAGAGAAGTTCCTTCAAGTGSAAACGACGCCCTGSAAGAGATCTTGAGAAGCA 1330
S I P E N N K E E L L O V E R R L E : L R S
S I P E N N K E E L L O V E R R L E : L R S

212/270

FIG. 15 CONTINUED.

Ava III
 Nsi I Xba I

AAGAATCATGCATGCTAATTCAGATAATATCCCAAAGAATCGAATTGCATTGTGTGATCCGTTTTTGC 1400
 K E S C I V : L D N I P K N R : A F V V S V F A
 K E S C I V : L D N I P K N R : A F V V S V F A
 EcoR V

AAATGTCCCACTTCAAAACAACGAAGGTCCATTGTAGTATGCACAGTCAACCGATATCAAAATCCCTGAG 1470
 V V P L G N N E G P F V V C T V N R Y G I P E
 V V P L G N N E G P F V V C T V N R Y G I P E
 CTTCAAAATTCACCAATTTCAAAATGTCAAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACC 1540
 L Q I H H N F K M S V M S N R L E G F : L R Y
 L Q I H H N F K M S V M S N R L E G F : L R Y
 TCCGACGACGGGCGGTAGAGGATGAGTATCGTCTAATCTGTACAGATGCCATCAGAGCTCTCAAAATCAT 1610
 L R R R A V E D E Y R L T V D M P S E L F K I :
 L R R R A V E D E Y R L T V D M P S E L F K I :
 EcoR I

TGACTCTTCCCAATAGCTCTTCAGGCGGTCAATAATTTTATGAGAAAACGAATTCCTGTGATGTGACA 1680
 D F F P : A L G A V N N F I E K T N S V D V T
 D F F P : A L G A V N N F I E K T N S V D V T
 Bam HI

GTTSGTCCAAGAGCATGCTTGAACGTGCTCTAATCTGTCGATGGATCCCGTGAATGGTTCATTCGATTGT 1750
 V G P R A C L N C P L T V D G S R E W F : R L
 V G P R A C L N C P L T V D G S R E W F : R L
 GGAATGAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAACCTTCGGTCGCTGCAC 1820
 V N E N F : P V L E R V A R D G K K T F G R C T
 V N E N F : P V L E R V A R D G K K T F G R C T
 Bam HI Tth I

TTCTTTCGAGGATCCACCGACATCGTCTCTAAAAAATGSCCGTGGTTCGATGGTGA AAAACCCGGAGAAT 1890
 S F E D P T C : V S K Y V D W F D S E N D E N
 S F E D P T C : V S K Y V D W F D S E N D E N

213/270

FIG. 15 CONTINUED.

Tth I

GTGC¹CAAACGTCTTCAACTCCAAGACCTCGTCCCTCAGCTGCCAACTCATCCCACAACACTTCAATC 1960

V L K R _ Q L Q D L V P S P A N S S S R Q H F N
 V L K R _ Q L Q D L V P S P A N S S S R Q H F N

Ava I
Xho I

CCCTCGASTCGTTGATCCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTC 2030

P L E S L I Q L H A T K H Q T I D N I
 P L E S L I Q L H A T K H Q T I D N I

Asp 718
Kpn I

TATCTTCTCTGCGCTCTCCCCGCTTTCCTTATCTTCGTACCGGTACCTGATGATTCCTCCCATTTTCCCC 2100

Ava I
Xma I
Sma I

CTTTTCCCCCAATTTCCGAGAACCTCTCTGTTCCTTGTTCCTAGTCTCTCCCGGTGCCGACGCCGAAG 2170

CGATTTAAAAACCTTTTCTTTCGGAACATTTCCCATTTGCTCATTAAATAGTCAAATTGAATAAACAGTG 2240

Dra II
Dra II
Pss I
Apa I
Pss I

TATGTACTTAAAAAAAAAAAAAAAAAAAAAAAAAGGGCCCTATTCTATAGTGTACCTAAATGCTAGA 2310

Bcl I

GCTCGCTGATCAGCTCGACTGTGCC¹TTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCGGTGCCTT 2380

CCTTGACCTGGGAAGGTGCCACTCCACTGTCTTTCCTAATAAAATGAGGAAATTSCATCGCATTTGCTT 2450

GASTAGGTGTCTATCTTATCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTTGGGAAGACAAT 2520

Pvu II

AGCAGGCATGCTGGGGATGGGTGGGCTCTATGGCTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGG 2590

GGTATGCCACGGCGCCCTGTAGCGGGGCAATTAAGCGGGCGGGTGTGGTGGTTACCGCAGCGTGACCGC 2660

214/270

FIG. 15 CONTINUED.

Nae I
TACACTTGCCAGCGCCCTAGCGCCGCTCCTTTTCGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGC 2730
TTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTAGTGCTTTACGGCACCTCGACC 2800

Dra III
CCAAAAAATTGATTAGGGTGATGGTTACGTTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCCCTTT 2870
GACGTTGGAGTCCCGTTCTTTAATAGTGGACTCTTGTTCGAACTGGAACAACACTCAACCTATCTCG 2940
GTCTATTCTTTTGAATTATAAGGGATTTTGGGGATTTCCGCCATTGGTTAAAAAATGAGCTGATTTAAC 3010
AAAAATTTAAGCGCAATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTTGAAAGTCCCCAGGCTCCCCA 3080

Ava III
Nsi I
GGCAGGCAGAAATATGCAAAGCATGCACTCAATTAGTCAGCAACCAGGTGTGGAAGTCCCAGGCTCC 3150

Ava III
Nsi I
CCAGCAGGCAGAAATATGCAAAGCATGCACTCAATTAGTCAGCAACCATAGTCCGCCCCCTAACTCCGC 3220

Nco I
CCATCCCCTCCCTTAAGTCCGCGCCAGTTCGCGCCATTCCTCCGCGCCATGGCTGACTAATTTTTTTTATTTA 3290

Stu I
Avr II
TGCAGAGGCGCGAGGCGGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTTA 3360

Ava I
Xma I
Sma I
Bcl I
GGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGAT 3430

Xma III
CGTTTCCATGATTGAACAAGATGGATTGCACGCGGGTCTCCGCGCGCTTGGGTGGAGAGGCTATTTCGG 3500

Nar I
Bbe I
CTATGACTGGGCACACAGACAAATCGCTTGGCTGATGCCGCGGTCTCCGCTGTGAGCGAGGGGCGC 3570

FIG. 15 CONTINUED.

Pst I	
CCGGTCTTTTSTCAAGACCGACCTGTCCGGTCCCTGAAAGAAGTGCAGGACGAGGCAGCGCGGCTAT	3640
Bal I Fsp I Pvu II Tth I	
CGTGGCAGCCACGAGGCGCTTCTTGCAGAGCTGTGCTCGACGTTGCACTGAAGCGGGAAGGGACTG	3710
GCTGCTATTGGCGAAGTGCCTGGGGCAGGATCTCTGTGATCTCACCCTCTCCAGCAGAAAGTATCC	3780
ATCATGGCTGATGCAATGCGGCGCTGTCATACGCTTGATCGCGCTACCTGCCCATTCGACCACCAAGCGA	3850
AACATCGCATCGAGCGAGCAGCTACTCGGATGGAAGCCGCTCTTGTGATCAGGATGATCTGACGAAGA	3920
BssH II	
GCATCAGGGCTCGCGCCAGCGGAGCTTTCGCCAGGCTCAAGGCGCGCATGCCGACGGCGAGGATCTC	3990
Nco I	
GTCGTGACCCATGGCGATGCCCTGCTGCGGAATATCATGCTGGGAAAAAGCGGCTTTCTGGATTTCATCG	4060
Nae I Rsr II	
ACTGTGGCGCGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGA	4130
GCTTGGCGGGAATGGCGTGAACCGTCTCTGTGCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATC	4200
Asu II	
GCCTTCATCGCCTTCTTGACGASTTCTTCTGAGCGGGACTCTGGGGTCAAAATGACCGACCAAGCGAC	4270
GCCCAACCTGCCATCAGGAGATTGGATTGACCGCCGCTCTATGAAAGGTTGGGCTTCGGAATCGTT	4340
Nae I	
TTCCGGGACGCCGGTGGATGATCTCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTGGCCACCCCAACT	4410
TGTTATTGAGGCTTATAATGCTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAGCATTTT	4480
Bsm I Sal I	
TCAGTGCATCTAGTGTGGTCTGCGAACTCATCAATGATCTTATCATGCTCTATACCGTCGACC	4550
CTAGCTAGAGCTTGGGCTAATCATGCTCATAGCTGTTCTCTGTGTGAATTTGTTATCGGCTCACAAATC	4620
CACACAACATACGAGCGCGGAAGCATGAAGTGTAAAGCCGGGGTGCCTAATGAGTGAAGCTAATCACAAT	4690

FIG. 15 CONTINUED.

Pvu II

AATTGCGTTGCGCTCAGTCTGCGGCTTTCCAGTCGGGAAACCTGTCTGCGGAGCTGCATTAAATGAATCGGC 4760
 CAACGCGCGGGGAGAGGCGCTTTTCCGCTATTGGGCGCTCTTCCGCTTCCCTCGCTCACTGACTCGCTGCGCT 4830
 CGGTGCTTCGGCTGCGGCGAGCGCTATCAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAGG 4900
 GGATAACGCAGGAAAGAACAATGTTAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGCCGCGCTG 4970
 CTGGCGTTTTCATAGGCTGCGGCGGCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAGAGGTGGC 5040
 GAAACCCGACAGGACTATTAAGATACCAGGCGTTTCCCGCTGGAAGCTGCGCTCGTGGCTCTCTCTTCC 5110
 GACCTGCGGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCA 5180

ApaL I

CGCTGTAGGTATCTCAGTCTGGTCTAGGTCGTTCTGCTCCAAGCTGGGCTGTGTCACGAACCCCGCTTC 5250
 AGCCCGACCGCTGCGGCTTATCGGCTAACTATCGTCTTGAGTCCAACCGGTAAGACACGACTTATCGCC 5320

AlwI

ACTGCGAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGATATGTAGGCGGTGCTACAGAGTCTTGAAG 5390
 TGGTGGCTTAACACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT 5460
 TCGGAAAAAGAGTTGGTAGCTCTTATCGGCAAAACAAACACCGCTGCTAGCGGTGCTTTTGTTTG 5530
 CAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTATCTCTTACGGGCTCGAC 5600

BspH I

GCTCAGTGGAAACGAAAACTCAGCTTAAGGGATTTTGGTCATGAGATTATCAAAAAAGGATCTTCACTAGA 5670
 TCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTAATATGAGTAAGCTTGGTCTGACAGTTA 5740
 CCAATGCTTAATCAGTGAGGCACTATCTCAGCGATCTGTCTATTTGCTCATCTATAGTTGCTGACTC 5810
 CCGTCTGTAGATAACTACATACGGGAGGGCTTACCATCTGGCGGCACTGCTTCAATGATACCGCGAG 5880
 ACCACGCTCACCAGCTCGAGTTTATCAGCAATAAACCAGCCAGCGGGAAGGCGGAGCGCAGAAAGTGG 5950
 TCTTCAACTTTATCGGCTCATATCAGTCTATTAATGTTGCGGGGAAAGCTTAAGTAAGTAGTTCCCA 6020

FIG. 15 CONTINUED.

Fsp I

GTTAATAGTTTGGSCAACSTTTTSCCAATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGG 6090
 CTTCAATCAGCTCCGCTTCCCAACGATCAAGGCGAGTTACAATGATCCCCCATGTTTGCACAAAAAGCGT 6160

Pvu I

TAGCTCCTTCGGTCCCTCCGATCGTTGTGCAAGTAAGTTGGCCGCGAGTGTATCACTCATGGTTATGGCA 6230

Sca I

GCACCTGCATAATTCTCTTACTTTCATGCCATCCSTAAGATGCTTTCTGTGACTGGTGAGTACTCAACCA 6300
 AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGCGTCAATACGGGATAATACCGC 6370
 GCCACATAGCAGAACTTTAAAGTGCTCATCATTTGSAACGCTTCTTCGSGGCGAAACTCTCAAGGATC 6440

ApaL I

TTACCGCTGTGAGATCCAGTTGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT 6510
 TCACCAGCGTTCCTGGGTGATCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG 6580

Ssp I

BspH I

GAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCAATTATCAGGSTTATTGTCTCATG 6650
 AGCGGATACATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCGCGGCACATTTCCCGGAAAAG 6720

Sal I

Bgl II

Sal I

TGCCACCTGACGTGACGATCGGGAGATCTCCGSAATCCCTATGGTCGACTCTCAGTACAATCTGCTCT 6790

AlwN I

GATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTTGTTGGAGGTGCTGAGTAGTSCGCGAGCA 6860
 AAATTTAAGCTACCAACAAGGCAAGGCTTGACCGACAATTCATGAAGAATCTGCTTAGGGTTAGGCGTT 6930

Nru I

Mlu I

Spe I

TGCGCTGCTTCCGATGTACCGCCAGATATACCGCTTGACATTTGATTTATTGACTAGTTATTAATAGTAA 7000
 TCAATTCGCGGTCTATTTAGTTATAGCCCATATATGAGTTCCGCGTTACATAACTTACCGTAAATGSCC 7070
 CGCGTGGCTGACCGGCCAAATGACCCCGGCCATGACGTCATTAATGACGTATGTTCCCTAGTAACGCC 7140

AATAGGGACTTTCCATTTGACGTCATAGGGTGGACTATTTACGGTAAATGCCCACTTGGCAGTACATCAA 7210

Nde I

GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGCTAAATGGCCCCCTGGCATTATGCCC 7280
 AGTACATGACC

7292

218/270

FIG. 16.

Sst I
 TATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCAC 70
 Ppa I Hind III Sst I Bam HI Spe I Xma III EcoR I
 TATAGGGAGACCCCAAGCTTGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTG 140
 R P P V C W N S
 Bgl II Afl III Mlu I
 CAGATCTTGGCTATCAAATTGGTGAACCTCGACGCGTCATTGGAGACTCCACAACCATGATAACCAAGCCA 210
 A D L G Y Q : S E L R R V : G D S T T M : T S H
 TCCAACCTGACATTCTTACTTCTCAACTACAATCCGAATGTTTCATGCACGGTGCCGCACAGAGTCGCTA 280
 P T D : L T S S T T I R M F M Y G A A Q S R V
 P T D : L T S S T T I R M F M Y G A A Q S R V
 GACAGTCTGGTCTTGATATGCTTCTTCCAAAGCAAATGATTCTCCAACCTCGTCAAGTCAATTTTGACAG 350
 D S L V L D M L L P K Q M : L Q L V K S : L T
 D S L V L D M L L P K Q M : L Q L V K S : L T
 Bbv II
 AGAGACGTCTGGTGTAGCTGGAGCAACTGGAATTGGAAGAGCAAAGTGGCGAAGACCCCTGGCTGCTTA 420
 E R R L V L A G A T G I G K S K L A K T L A A Y
 E R R L V L A G A T G I G K S K L A K T L A A Y
 Asu II Bsm I
 TGTATCTATTGGAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCATTCTGAAAACAATAAAGAA 490
 V S I R T V Q S E D S : V N : S I P E N N K E
 V S I R T N Q S E D S : V N : S I P E N N K E
 Xmn I Bgl II Ava III Nsi I Xba I
 GAATTGCTTCAAGTGGAAACGACGCTGGAAAAGATCTTGAGAAGCAAAGAATCATGCATCGTAATTCTAG 560
 E L L O V E R R L E K I L R S K E S C : V I L
 E L L O V E R R L E K I L R S K E S C I V I L
 ATAATATCCCAAAGAATCGAATTCGATTGTGTATCCGTTTTSCAAATGTCCCACTTCAAAACAACGA 630
 D N I P : N R : A F V V S V F A N V P L Q N N E
 D N I P : N R : A F V V S V F A N V P L Q N N E
 EcoR V
 AGGTCCATTGTAGTATGCACAGTCAACCGATATCAAAATCCCTGAGCTTCAAAATTCACCACAATTTCAAA 700
 S P F V V C : V N P Y C I P E L O I H N F K
 G P F V V C : V N R Y C I P E L O I H N F K

FIG. 16 CONTINUED.

ATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCCTACGTTACCTCCGACGACGGGCGGTAGAGGATG 770
M S V M S N R L E G F I L R Y L R R R A V E D
M S V M S N R L E G F I L R Y L R R R A V E D
Sst I
AGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACTTCTTCCCAATAGCTCTTCA 840
E Y R L T V Q M P S E L F K I I D F F P I A L Q
E Y R L T V Q M P S E L F K I I D F F P I A L Q
EcoR I
GGCCGTCAATAATTTTATTGAGAAAACGAATTCGTGATGTGACAGTTGGTCCAAGAGCATGCTTGAAC 910
A V N N F I E K T N S V D V T V G P R A C L N
A V N N F I E K T N S V D V T V G P R A C L N
Bam HI
TGTCCTCTAACTGTGATGGATCCCGTGAATGGTTCATTGATGTTGGAATGAGAACCTCATTCCATATT 980
C P L T V D G S R E W F I R L V N E N F I P Y
C P L T V D G S R E W F I R L V N E N F I P Y
Afl III Bam HI Tth I
TGGAACGTGTTGCTAGAGATGGCAAAAAAACCTTCGGTCTGCACTTCCTTCGAGGATCCACCGACAT 1050
L E R V A R D S K K T F G R C T S F E D P T D I
L E R V A R D S K K T F G R C T S F E D P T D I
Bbv II
CGTCTCTAAAAATGGCCGTGCTTCGATGGTGAAAAACCCGAGAATGTGCTCAAACGTCTTCAACTCCAA 1120
V S K K V P W F D G E N P E N V L K R L Q L Q
V S K K V P W F D G E N P E N V L K R L Q L Q
Tth I Xho I
GACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCTCGAGTCGTTGATCCAATTGC 1190
D L V P S P A N S S R Q H F N P L E S L I Q L
D L V P S P A N S S R Q H F N P L E S L I Q L
Bbv II
ATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATCTTCTCTCGCCTCTCCCCCG 1260
H A T K H O T I D N I
H A T K H O T I D N I
CTTTCCTTATCTTCTACCGTACCTGATGATTCGCCATTTTCCCCCTTTTCCCCCAATTTCCCGAGAAC 1330
Xma I
Sma I
CTCCTGTTCCCTTTTGTCTAGTCTCTCCGSGTGCCGACGCCGAAGCGATTTAAAAACCTTTTCTTTCC 1400
Xmn I
GAAACATTTCCCAATTGCTCATTAATAGTCAAAATTAATAAACAGTGTATGTACTTAAAAAAAAAAAAAAAA 1470

FIG. 16 CONTINUED.

Sst I Bcl I
 AAAAAAAAAAAGGGCCCTATTCATAGTGTACCTAAATGCTAGAGCTCGCTGATCAGCCTCGACTGTG 1540
 CCTTCTAGTTGCCASCCATCTGTGTTTGCCCTCCCCCTGCTTCTTTGACCTGGAAGGTGCCACTC 1610
 CCACTGTCTCTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGG 1660
 Bbv II
 GGGTGGGGTGGGSCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTG 1750
 GGCTCTATGGCTTCTGAGGCGGAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCG 1820
 GCGCATTAAAGCGCGCGGCTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCC 1890
 CGCTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGCTTTCCCGCTCAAGCTCTAAATCGG 1960
 GGCATCCCTTTAGGCTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAAC TTGATTAGGGTGTATG 2030
 Dra III
 GTTCACGTATGSSCCATGSCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAA 2100
 TAGTGGACTCTTTTCCAAACGGAACAACACTCAACCTATCTCGGTCTATTCTTTTGATTATAAGGG 2170
 Xmn I
 ATTTTGGGGATTTCGGCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCT 2240
 GTGGAATGTGTTTCAGTTAGGGTGTGGAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCAT 2310
 Ava III
 Nsi I
 GCATCTCAATTATGTCAGCAACCAGGTGTGSAAGTCCCCAGGC*CCCCAGGCAGGCAGAAGTATGCAAAGC 2360
 Ava III
 Nsi I
 ATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAACTCCGCCCA 2450
 GTTCCGCCCA*TTCTCCGCCCATGGCTGACTAATTTTTTTATTTATGCAAGGCCGAGGCCGCCCTCTGC 2520
 Stu I
 Xma I
 Sma I
 CTCTGAGCTATTCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCC TAGGCTTTTGCAAAAAGCTCCCGGGA 2590
 Bcl I
 GCTTGATATCTATTCTCGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATG 2660
 Xma III
 GATTGCACGCAGGTTCTCCGGCCGCTTGGSTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGACAAT 2730
 Nar I
 Bbe I
 CGGCTGCTCTGATGCGCGCTGTTCCGGCTGTGAGCGCAGGGGCTCCCGCTTCTTTTGTCAAGACCGAC 2800
 CTGTCCGGTGCCTGAA*GAAC*GCAGGACGAGGCAGCGCGCTATCGTGGTGGCCACGACGGCGGTTG 2870
 Fsp I
 Tth I
 CTTGCGCAGCTGTGCTGACGCTGTGCTACTGAAGCGGGAAGGGAC TGGCTGCTATTGGCGAAGTGCGGG 2940
 GCAGGATCTCTGTCACTCACTTGTCTCTCCGAGAAAGTATCCATCATGCTGATGCAATGCGGCGG 3010

FIG. 16 CONTINUED.

CTGCATACGCTTATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA 3080
CTCGGATGGAAGCCGCTTGTCSATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGA 3150

BssH II

ACTGTTCCGCGAGGCTCAAGGCGCSCATGCCCGACGGCGAGSATCTCGTCGTGACCCATGGCGATGCCTGC 3220

Rsr II

TTGCCGAATAATCATSGTGSAATATGGCCGCTTTTCTGGATTTCATCGACTGTGGCCGGCTGGGTGTGGCGG 3290
ACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCG 3360
CTTCCTCGTGCTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAG 3430

Asu II

TTCTTCTGAGCGSSACTCTGGGGTTGAAATGACCGACCAAGCGAGCGCCCAACCTGCCATCAGGATTT 3500
CGATTCCACCGCCGCTTCTATGAAGGTTGGGCTTCGGAATCGTTTCCGGGACGCCGGCTGATGATC 3570
TCCAGCGCGGGATCTCATGCTGAGTTCTTCGCCACCCCACTTGTATTGTCAGCTTATAATGGTT 3640

Bsm I

ACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTCACTGCAATTCAGTTGTGGTTT 3710

Sna I

GTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCCGACCTCTAGCTAGAGCTTGGCGTAATCA 3780
TGGTCATAGCTGTTTCTGTGTGAAATTTTATCCGCTCACAAATTCACACAACATACGAGCCGGAAGCA 3850
TAAAGTGTAAGGCTGGGGTGCTTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC 3920

Sbo I

TTTCAGTCCGGAAACCTGTCTGTCCAGCTGCATTAATGAATCGSCCAACGCGCGGGAGAGGCGGTTTG 3990
CGTATTGGGCGCTCTTCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTGCGCTGCGCGGAGCGG 4060

AII III

TATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTG 4130
AGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGC 4200
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTTGGCGAAACCCGACAGGACTATAAAGAT 4270
ACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCCTCTCTGTTCCGACCTTGCCGCTTACCGGATACCT 4340
GTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTGGGTG 4410

ApaL I

TAGGTCGTTCCCTCCAAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCCTTATCCG 4480

AlwI I

GTAACTATCTCTTSGAGTCCAAACCCGGTAAGACACGACTTATGCCACTGGCAGCAGCCACTGGTAACAG 4550
GATTAGCAGAGCAGGTATGTAAGCGGTGCTACAGAGTTCTTGAAGTGGTGGCTTAACACGGCTACACT 4620
AGAAGSACAGTATTGGTATCTCCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGAGTTGGTAGCTCTT 4690
GATCCGGCAAAACCAACCAACCGCTGATAGCGGTGTTTCTTCTGCAAGCAGCAGATTACGCGCAGAAA 4760
AAAAGGATCTCAAGAGATCTTATCTTTCTACGGGCTCTGACGCTCAGTGGAACGAAAACTCACGT 4830

BspH I

TAAAGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAAAGAT 4900

FIG. 16 CONTINUED.

TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCASTGAGGCACC 4970
TATCTCAGCGATCTGTCTATTCTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA 5040
Ppa I
CGGGAGGGCTTACCATCTGSCCCAGTCTGCAATGATACCGGAGACCCACGCTCACCGGCTCCAGATT 5110
TATCAGCAATAAACCAGCCAGCCGSAAGG3CCGAGCGCAGAAGTGGTCTTGAACTTTATCCGCTCCAT 5180
Fsp I
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCGAGTTAATAGTTTGGCGAACGTTGTT 5250
GCCATTGCTACAGGCATCTGGTGTAC3CTCGTCTGTTGGTATGGCTTCATTCAGCTCCGGTTCCCAAC 5320
Pvu I
GATCAAGGCGAGTTACATGATCCCCAT3TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGT 5390
TGTCAGAAAGTAAGTTGGCCGCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC 5460
Sca I Sbo I
ATGCCATCCGTAAGATGCTTTCTGTGACTSGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC 5530
GGC3ACCGAGTT3CTCTTGC003CGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGT 5600
Xmn I
GCTCATCATTGGAAAACGTTCTTCCGGGCG3AAAACCTCTCAAGGATCTTACC3CTGTTGAGATCCAGTTCC 5670
EcoK ApaL I
ATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAA 5740
AAACAGGAAGGCAAAATGCCGCAAAAAAG33AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT 5810
Ssp I BspH I
CCTTTTCAATATTATGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATT 5880
TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCG3AAAAGTGCCACCTGACGTGACGGATCGG 5950
Bgl II AlwN I
GAGATCTCCCGATCCCTATGGTCSACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT 6020
CTGCTCCCTGCTTGTGTGTTGCAAGTCCGTSAGTAGTGC3GAGCAAAATTTAAGCTACAACAAGGCAAG 6090
Nru I
GCTTGACCGACAATTGCATGAAGAATCT3CTTAGGGTTAGGCGTTTTTGGCTGCTTCCGATGTACGGCC 6160
Alu III Mlu I Spe I
CAGATATACCGGTTGACATTGATTATGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTAT 6230
AGCCCATATATGGAGTTC3CGCTTACATACCTTACGGTAAATGGCCCGCTG6CTGACCGCCCAACGACC 6300
CCCGCCCATTGAC3TCAATAATGACGTAT3TCCCATAGTAACGCCAATAGGGACTTTCCATTGAC3TCA 6370
Nde I
ATGGTGGACTATTTACGGTAAACT3CC3ACTTGG3AGTACATCAAGTGATCATATGCCAAGTAC3CC 6440
CCTATTGAC3TCAATGAC3TAAAT3GGCC3GCTG3CATTAAGCCAGTACATGACCTTATGG3ACTTTC 6510
SnaB I
CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTTGGCAGTACATCAA 6580
TGGCGTGGATAGCGGTTTGACTCACGGG3ATTTC3AAGTCTCCACCCCATGACGTCAATGGGAGTTTG 6650
TTTTGGCACCAAAATCAACGGGACTTTCC3AAAATGTCTGAACAAC3CGCCCATTGACGCAAAATGGGCG 6720
GTAGGCGTGTACGGTGGGAGGTCTA 6745

FIG. 17.

GGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATGTTGCCGGGAAGCTAGAGTAAGTAGTTCGC 70
Fsp I
CAGTTAATAGTTTSCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTAT 140
GGCTTCATTGAGTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCG 210
Pvu I
GTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCGAGTGTATCACTCATGGTTATGG 280
Sca I
CAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTCTGTGACTGGTGAGTACTCAAC 350
CAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACC 420
GCSCCAGATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGA 490
ApaL I
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCAACTGATCTTCAGCATCTTTTAC 560
TTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCSAAAAAAGGSAATAAGGGCGACA 630
Ssp I BspH I
CGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGCTCTCA 700
TGAGCGGATACATATTTGAATGTATTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCCGAAA 770
Sal I Bgl II Sal I
AGTGCCACCTGACGTGACGGATCGGGAGATCTCCGATCCCCATGGTGGACTCTCAGTACAATCTGCT 840
AlwI
CTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTGCGTGTAGTAGTGGCGGAG 910
CAAAA'TTAAGCTACAACAAGGC AAGGCTTSACCSACAATTGCATGAAGAATCTGCTTAGGGTAGGCGT 980
Nru I Mlu I Spe I
TTTSCGCTGC'TTCGATGTACGSGCCAGATATACSGCTTGACATTGATTATTGACTAGTTATTAATAGT 1050
AATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTACATAACTACGGTAAATGG 1120

FIG. 17 CONTINUED.

CCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACG 1190
CCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATC 1260
Nde I
AAGTGATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGCTAAATGGCCCGCCTGGCATTATGC 1330
SnaB I Nco I
CCASTACATGACCTTATGGGACTTTCTTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG 1400
GTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCAGGGGGATTTCGAAGTCTCC 1470
ACCCCATGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTGAACAA 1540
CTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGG 1610
Hind III
CTAACTAGAGAACCCTACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCCAAGCTT 1680
L E N D L L T G L S K L I R L T : G R P K L
Asp 718 Bam HI
Kpn I Spe I Xma III BstX I EcoR I Pst I
GGTACCGAGCTCGGATCCACTAGTAACGCGCGCCAGTGTGCTGGAATCTGCAGATCGCCGCTGCCACCT 1750
G T E L S S T S N G R Q C A G : L Q I A A A T
Ava I
Xma I
Sma I
Pvu II
CAACCTTCGGACAACATTGCTTAAGATCCCGGGATACTCATCTATTCTCCACACTTATCAGTGTACGC 1820
S T F G Q H S L R S P G Y S S Y S P H L S V S A
Spe I
Sal I
TGATAAGGACACAATGTCTATGCACTCACAGACTAGTCCGACGACCTTCTTCACAAAAACCAAGCTATTCA 1890
D Y D T M S M F S C T S R R P S S Q K P E / S
M E M F S Q T S R R P E S Q K P E / S
GGCCAAATTCATTGACTTGAATGCTAAATGCCACCTTCAAGAGTTGACATCCACCGAGCACAGAAATGGCGG 1960
S Q F H S L C R K C H L Q E F T S T E F R V A
S Q F H S L C R K C H L Q E F T S T E F R V A

FIG. 17 CONTINUED.

Ava I Bam HI
CTCTCTT¹GAGCCCGAGACGGGTGCCAACTCGATSTCGAAATATGATTCTTCAGGATCCTACTCGGC3CG 2030
A L L S P R R V P N S M S K Y D S S G S Y S A R
A L L S P R R V P N S M S K Y D S S G S Y S A R
Ava I
TCCCGAGGTGGAAGCTCTACTGSTATCTATGGAGAGACGTTCCAAC²TGCACAGACTATCCGATGAAAAA 2100
S R G G S S S T G I Y G E T F Q L H R L S D E K
S R G G S S S T G I Y G E T F Q L H R L S D E K
Bam HI Nde I
TCCCCGACATTCCTGCCAAAAGTGAAGATGGGATCCCAACTATCACTGGGTAGCAGCAGACAGCATATGGAT 2170
S P A H S A K S E M G S Q L S L A S T T A Y G
S P A H S A K S E M G S Q L S L A S T T A Y G
Sal I
CTCTCAATGAGAAGTACGAACATGCTATTCGGGACATGGCACGTGACTTGGAGTGTACAAGAACA³CTGT 2240
S L N E K Y E H A I R D M A R D L E C Y K N T V
S L N E K Y E H A I R D M A R D L E C Y K N T V
Hind III
CGACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTGTTTGATCTTTT⁴TGAGCAAAAGCTTAGAAAA 2310
D S L T K K Q E N Y G A L F D L F E Q K L R K
D S L T K K Q E N Y G A L F D L F E Q K L R K
Cla I
CTCACTCAACACATTGATCGATCCAACTTGAAGCCTGAAGAGGCAATACGATTCAAGGCAGGACATTGCTC 2380
L T Q H I D R S N L K P E E A I R F R Q D I A
L T Q H I D R S N L K P E E A I R F R Q D I A
ATTGAGGGATATTAGCAATCATCTTGCATCCAACTCAGCTCATGCTAACGAAGGGCTGGTGAGCTTCT 2450
H L P D I S N H L A S N S A H A N E G A G E L L
H L P D I S N H L A S N S A H A N E G A G E L L
Cla I Cla I
TCTCAACCATCTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTGGTGGAAAAAGCAGCAAG 2520
R C P S L E S V A S H R S S M S S S S K S S K
R C P S L E S V A S H R S S M S S S S K S S K

FIG. 17 CONTINUED.

Bam HI

CAGGAGAAGATCAGCTTGAGCTGCTTTGGCAAGAACAAGAAGAGCTGGATCCGCTCCTCACTCTCCAAGT 2590

Q E K I S L S S F G K N K K S W I R S S L S K
Q E K I S L S S F G K N K K S W I R S S L S K

Nde I BspM II

TCACCAAGAAGAAGAACAAGAACTACGACGAAGCACATATGCCATCAATTTCCGGATCTCAAGGAAGCTC 2660

F T K K K N K N Y D E A H M P S I S G S G G T L
F T K K K N K N Y D E A H M P S I S G S G G T L

ApaL I

TCACAACATTGATGTGATTSAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACTTACGAAGTCCGC 2720

D N : D V I E L K Q E L K E R D S A L Y E V R
D N : D V I E L K Q E L K E R D S A L Y E V R

CTTGACAATCTGGATCGTGCCCGCAAGTTGATGTTCTGAGGGAGACAGTGAACAAGTTGAAAACCSAGA 2800

L D N L D R A R E V D V L R E T V N K L K T E
L D N L D R A R E V D V L R E T V N K L K T E

ACAAGCAATTAAAGAAAGAAGTGGACAAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCGCGCCTC 2870

N K Q L < K E V D K L T N G P A T R A S S S R A S
N K Q L < K E V D K L T N G P A T R A S S S R A S

AATTCAGTTATCTACGACGATGAGCATGTCTATGATGCAACGCTAGCATACATCAGCTAGTCAATCT 2940

I P V I Y D D E H V Y D A A C S S T S A S Q S
I P V I Y D D E H V Y D A A C S S T S A S Q S

Asu II

TCGAAACGATCCTCTGGCTGCAACTCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTT 3010

S K R S S G C N S I K V T V N V D I A G E I S
S K R S S G C N S I K V T V N V D I A G E I S

Pvu I Hpa I EcoR V

CGATCGTTAACCCTGACAAAGAGATAATCGTAGGATATCTTGCCATGTCACCCAGTCAGTCAATGCTGGAA 3080

S : I N P D K E : I V G Y L A M S T S D S C L K
S : I N P D K E : I V G Y L A M S T S D S C L K

AGACATTGATGTTCTATTCTTAGSACTATTTGAAGTCTACCTATCCAGAAATTGATGTGGAGCATCAACT 3150

D : D V S I L S L F E V Y L S R : D V E H C L
D : D V S I L S L F E V Y L S R : D V E H C L

Cla I
 GGAATCGATGCTCGTGATTCTATCCTTGGCTATCAAATTGGTGAACCTCGACGCGTCATTGGAGACTCCA
 3220
 G I D A R D S : L G Y O I G E L R R V I G D S
 G I D A R D S : L G Y Q I G E L R R V : G D S
 CAACCATGATAACCAGCCATCCAACCTGACATTCTTACTTCTCAACTACAATCCGAATGTTTCATGCACGG
 3290
 T T M : T S H P T D I L T S S T T I R M F M H G
 T T M : T S H P T D I L T S S T T I R M F M H G
 TGCCGCACAGAGTCGCGTAGACAGCTCTGGTCCTTGATATGCTTCTTCAAAGCAAATGATTCTCCAACCT
 3360
 A A Q S R V D S - V L D M L L P K O M I L O L
 A A Q S R V D S - V L D M L L P K O M : L O L
 GTCAAGTCAATTTTGACAGAGAGACGCTCTGGTGTAGCTGGAGCAACTGGAATTGSAAGAGCAAACCTGG
 3430
 V K S : L T E R R L V L A G A T G : G K S K L
 V K S : L T E R R L V L A G A T G : G K S K L
 Asu II
 Bsm I
 CGAAGACCCTGGCTGCTTATGTATCTATTGCAACAAATCAATCCGAAGATAGTATTGTTAATATCAGCAT
 3500
 A K T L A A Y V S I R T N O S E D S : V N I S :
 A K T L A A Y V S I R T N O S E D S : V N I S :
 Bgl II
 CCTGAAAACAATAAAGAAGAATTSCCTTCAAGTGAACGACGCC TGGAAAAGATCTTGAGAAGCAAAGAA
 3570
 P E N N K E E L L O V E R R L E K I L R S K E
 P E N N K E E L L O V E R R L E K I L R S K E
 Ava III
 Nsi I
 Xba I
 CATGCATCGTAATTCTAGATAATATCCCAAAGAATCGAATTGCATTGTTGTATCCGTTTTTGCAAATG
 3640
 S C I V I L C V : P K Y R I A F V V S V F A N
 S C I V I L C V : P K N R I A F V V S V F A N
 EcoR V
 CCACTTCAAAACAACGAAGGTCATTTGTAGTATGCACAGTCAACCGATATCAAATCCCTGAGCTTCA
 3710
 P L O N N E S P F V V C T V N R Y C I P E L O
 P L O N N E S P F V V C T V N R Y C I P E L O
 ATTCAACCAATTTCAAATGTCAGTAATGTCGAATCGTCTCGAAGGATTCATCTACGTTACCTCCGA
 3780
 I H H V F K M S V M S N R L E G F I L R Y L R
 I H H V F K M S V M S N R L E G F I L R Y L R

FIG. 17 CONTINUED

CGACGGGCGGTAGAGGATGAGTATCGTCTAACTGTACAGATGCCATCAGAGCTCTTCAAAATCATTGACT 3850
R R A V E D E Y R L T V Q M P S E L F K I I D
R R A V E D E Y R L T V Q M P S E L F K I I D
EcoR I
TCTTCCAATAGCTCTTCAGGCCGTCAATAATTTATTGAGAAAACGAATTCTGTTGATGTGACAGTTGG 3920
F F P : A L C A V N N F I E K T N S V D V T V G
F F P : A L C A V N N F I E K T N S V D V T V G
Bam HI
TCCAAGAGCATGCTTGAAGTGTCTCTAACTGTCSATGGATCCCGTGAATGGTTCATTGCGATTGTGGAAT 3990
P R A C L V C P L T V D G S R E V F I R L V N
P R A C L V C P L T V D G S R E V F I R L V N
GAGAACTTCATTCCATATTTGGAACGTGTTGCTAGAGATGGCAAAAAACCTTCGGTCCGCTGCACTTCTC 4060
E N F : P Y L E R V A R D G K K T F G R C T S
E N F : P Y L E R V A R D G K K T F G R C T S
Bam HI Tth I
TCGAGGATCCACCGACATCGTCTCTAAAAAATGGCCGTGGTTCGATGGTGAAAACCCGAGAAATGTGCT 4130
F E D P T D : V S K K W P W F D G E N P E N V L
F E D P T D : V S K K W P W F D G E N P E N V L
Tth I Ava I Xho I
CAAACGTCTTCAACTCCAAGACCTCGTCCCGTCACCTGCCAACTCATCCCGACAACACTTCAATCCCTC 4200
K R L Q L Q D L V P S P A N S S R Q H F N P L
K R L Q L Q D L V P S P A N S S R Q H F N P L
GAGTCGTTGATCCAATTGCATGCTACCAAGCATCAGACCATCGACAACATTTGAACAGAAGACTCTAATC 4270
E S L : Q L H A T K H O T I D N I
E S L : Q L H A T K H O T I D N I
Asp 718 Kpn I
TTCTCTCGCTCTCCCCCGCTTCTTATCTTCGTACCGGTACCTGATGATTCCCATTTTCCCCCTTT 4340
Ava I Xma I Sma I
CCCCCAATTCCGAGAAGCTCTGTTCCTTTGTTCCTAGTCTCCCGGGTGGCGACGCCGAAGCGATT 4410

TAAAAACCTTTTCTTTCCGAAACATTTCCCATTCGTCATTAAATAGTCAAATGAATAAACAGTGTATGT 4460

Dra II
Dra II
Pss I
Apa I
Pss I

ACTTAAAAAAAAAAAAAAAAAAAAAAAAAGGGCCCTATTCTATAGTGCACCTAAATGCTAGAGCTCG 4550

Bcl I

CTGATCAGCCTCGACTGTGCCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCGCCCGTGCCTTCCTTG 4620

ACCCTGGAAGGTGCCACTGCCACTGTCTTTCTTAATAAAATGAGGAAATGTCATCGCATGTCTGAGTA 4690

GGTGTCAATCTATTCTGGGGGGGGGGGGGGGGGAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG 4760

Pvu II

GCATGCTGGGGATGCGGTGGGTCTATGGCTCTGAGGCGGAAAGAACCAGCTGGGGGCTCTAGGGGGTAT 4830

CCCCACGCGCCCTGTAGCGGGCATTAAAGCGGGCGGGGTGGGTGGTTACGCGCAGCGTGACCGCTACAC 4900

Nae I

TGCCAGCGCCCTAGCGCGCGCTCTCTTGGCTTTCTTCCCTTCTTTCTCGCCACGTTGCGCGGCTTTCC 4970

CCGTCAAGCTC~AAATCGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAA 5040

Dra III

AAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGT 5110

TGGAGTCCACGTCTTTAATAGTGGACTCTGTGTCGAAACTGGAACAACACTCAACCCATCTCGGTC TA 5180

TTCTTTTGATTTATAAGGGATTTGGGGATTTGGGCTATTGGTTAAAAATGAGCTGATTTAACAAAAA 5250

TTAACGCGAA~TAATTCGTGGAAATGTGTGTCAGT~AGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAG 5320

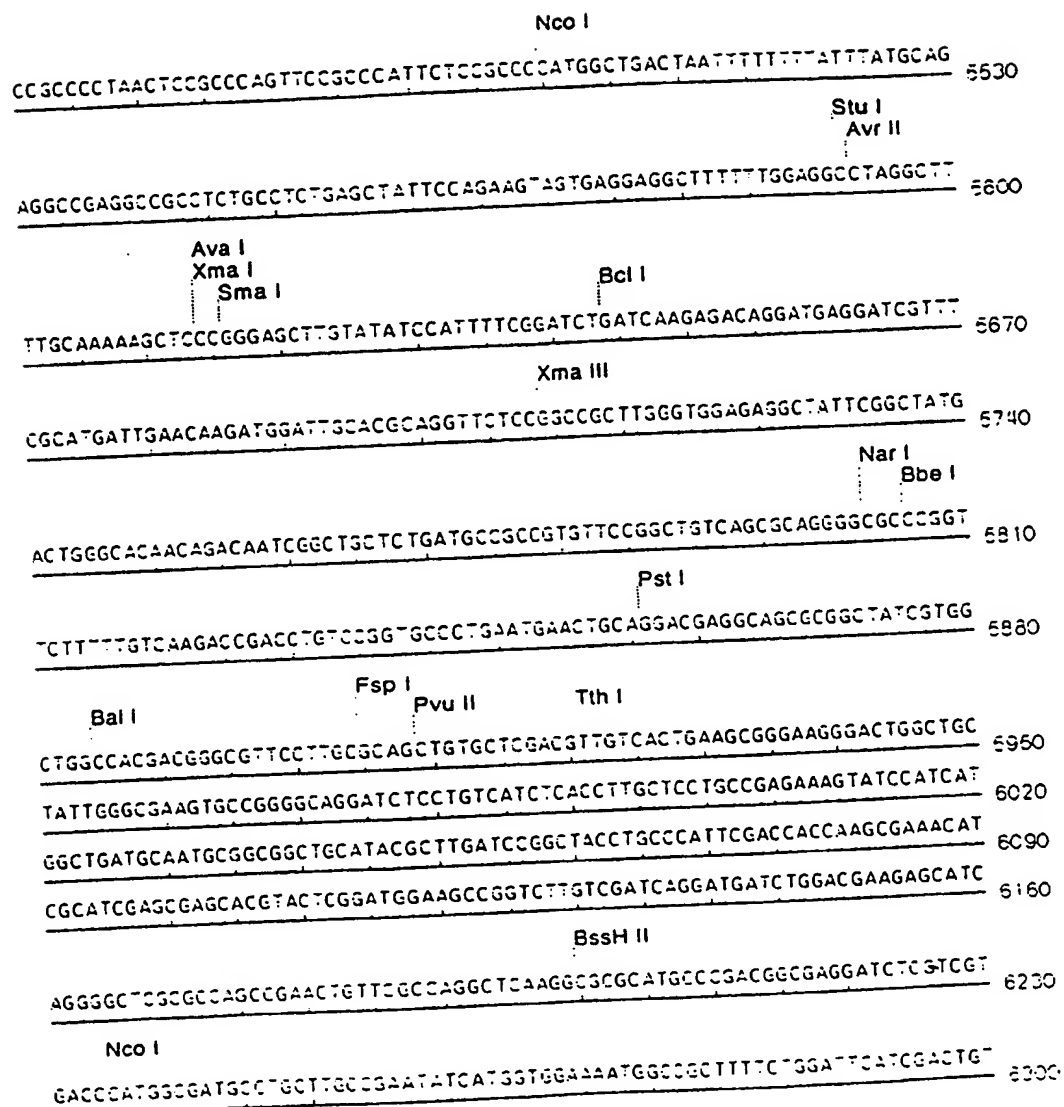
Ava III
Nsi I

GCAGAAGTATGCAAAAGCATGCA~TCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGC~TCCCCAGC 5390

Ava III
Nsi I

AGGCAGAAGTATGCAAAAGCATGCA~TCAATTAGTCAGCAACCATAGTCCCCCCCC~TAACTCGGCCCA~TC 5460

FIG. 17 CONTINUED.



Nae I **Rsr II**

GGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTG 6370
GGGGCAATGGGCTGACCGCTTCTCGTGTCTTACGGTATCGCCGCTCCCSATTCCAGCGCATCGCCTT 6440

Asu II

CTATCGCCTTCTTGACGAGTTCCTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCA 6510
ACCTGCCATCACGAGATTTCGATTCCACCSCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCGG 6580

Nae I

GGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTCTTCGCCACCCCAACTTGTTT 6650
ATTGCGAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTCTTTTCAC 6720

Bsm I **Sal I**

TGCATTCTAGTTGTGGTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAG 6790
CTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCGTGTGTGAAATTGTATCCGCTCACAATTCACAC 6860
AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCTTAATGAGTGAGCTAACTCATTAAATTG 6930

Pvu II

CGTTGCGCTCACTGCCCGCTTTCAGTCCGGAAACCTGTCTGCCAGCTGCATTAATGAATCGGCCAACG 7000
CGCGGGGAGAGGCGGTTTGGGTATTGGGCGCTCTTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTC 7070
GTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTATCCACAGAATCAGGGGATA 7140
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGSCCGCTTGCTGGC 7210
GTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC 7280
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTSCGCTCTCTGTTCGACCC 7350
TGGCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAGCTG 7420

ApaL I

TAGGTATCTCAGTTGGGTGAGGTCTTCCGTCGAAGCTGGGCTGTGTGACGAACCCCCCTCAGCCC 7490
GACCGCTGGCGCTTATCCGCTAACTATCGTCTTGAGTCCAAACCGGTAAAGACACGCTTATCGCCACTGG 7560

FIG. 17 CONTINUED.

AlwI

CASCAGCCACTGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTG 7620
GCCTAAC TACGGCTACACTAGAAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA 7700
AAAAGAGTTGGTASCTCTTGAATCCGCAAAACAAACCACCGCTGGTASCGGTGGTTTTTTGTTTGCAAGC 7770
AGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA 7840

BspH I

GTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT 7910
TTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT 7980
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCAATCCATAGTTGCCGACTCCCGT 8050
CGTGTAGATAAC TACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCA 8120
CGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT 8183



Figure 18 : Phase contrast images of MCF-7 cells transfected with pCB201 (upper) compared to mock (control) transfected MCF-7 cells (bottom).

The control cells are spread out on the tissue culture plastic and exhibit few filopodia outgrowths. The transfected cells appear smaller because they are slightly rounded up and have multiple filopodia outgrowths (arrowhead) per cell.

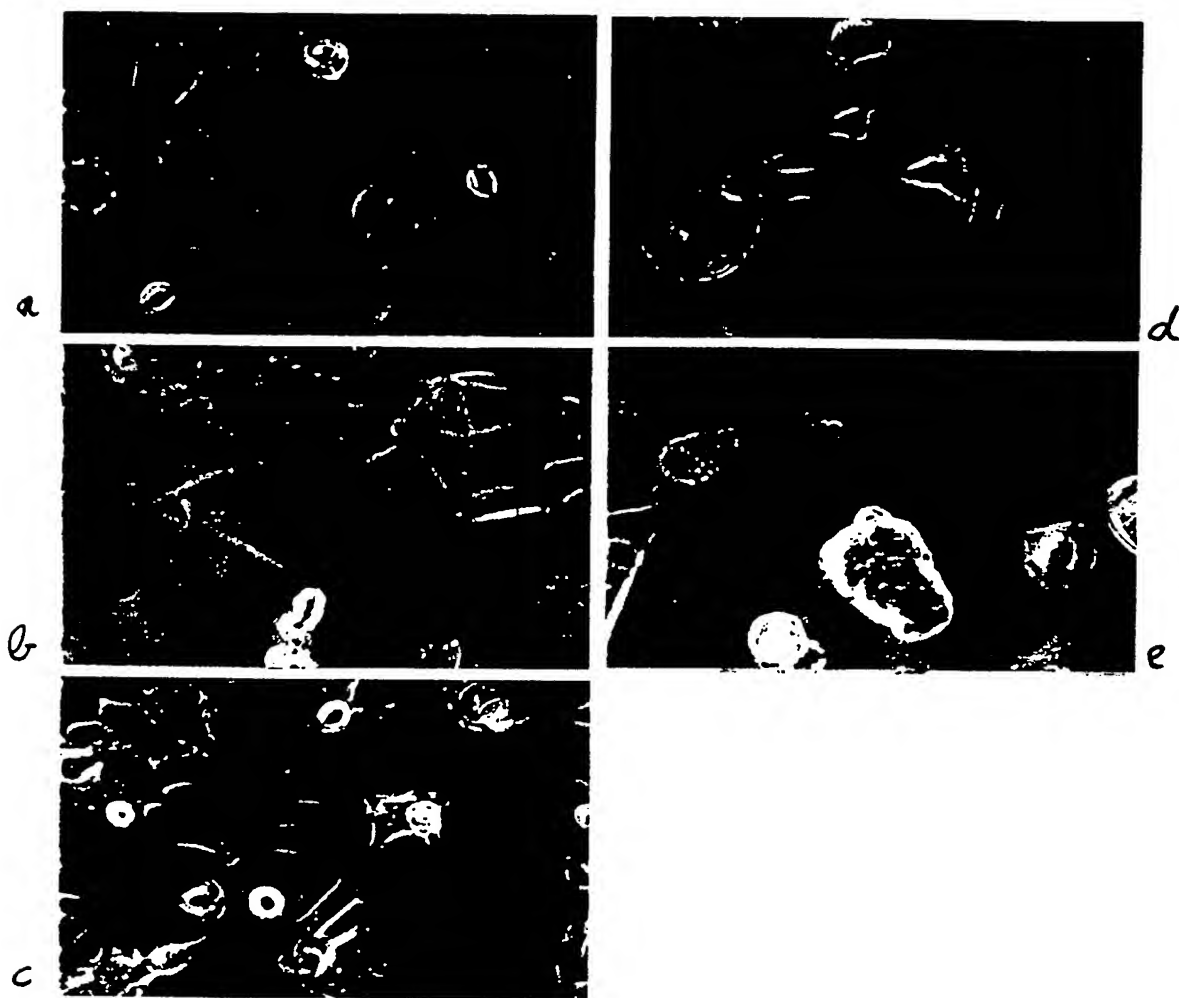


Figure 19 : Phase contrast image of MCF-7 cells, transfected with pcDNA3 (19a), pCDU4 (19b), pCDU3 (19c) pCDU2 (19d) and pTB72 (19e).

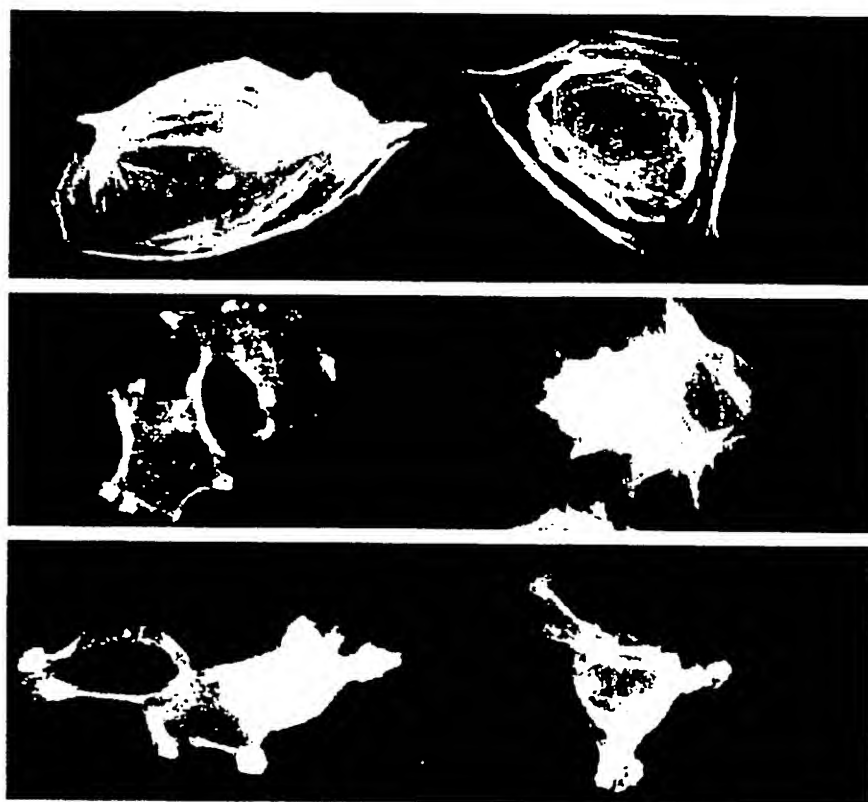


Figure 20 : F- actin pattern (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3.LacZ (top panel) and with pCB201 (middle and lower panel).

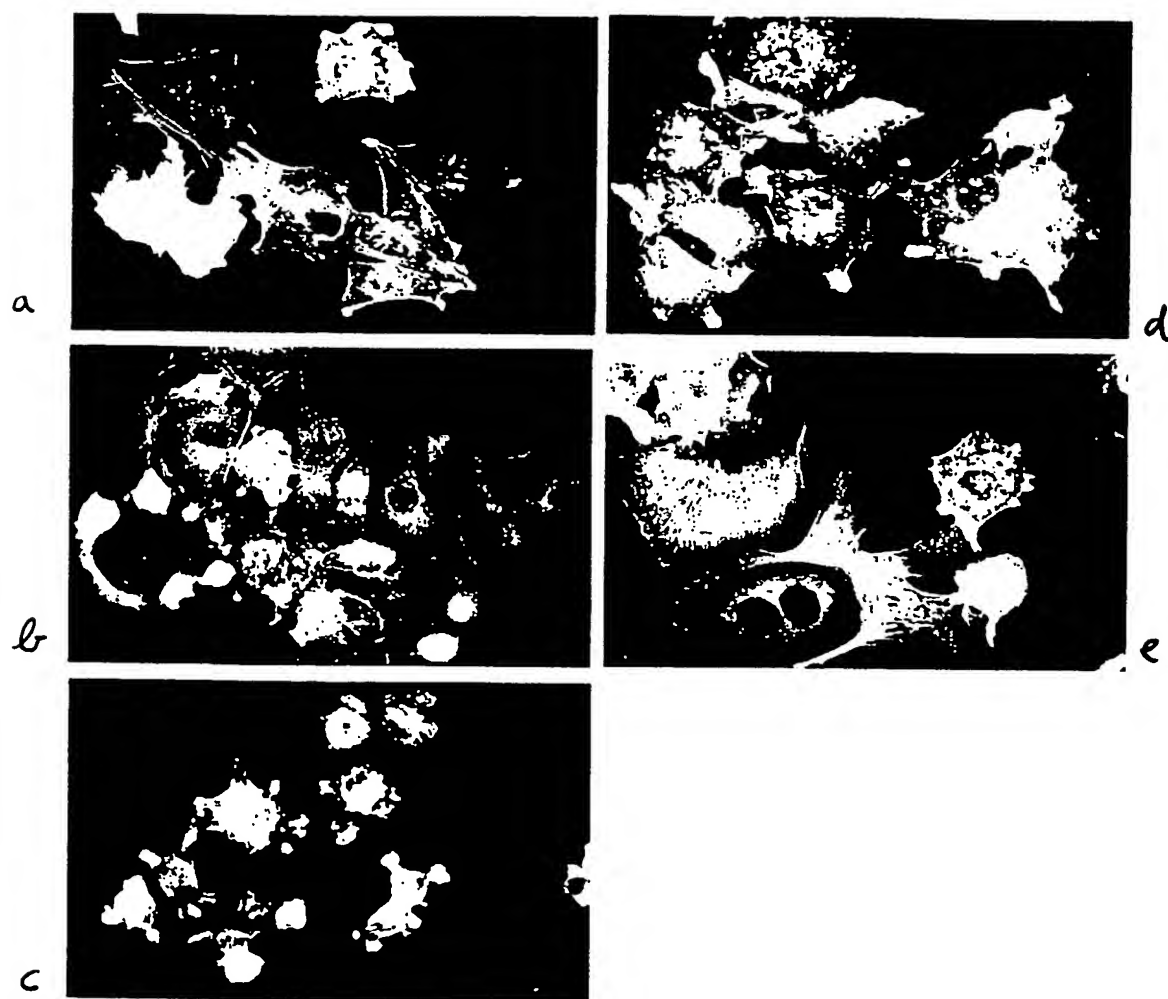


Figure 21 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of MCF-7 cells transfected with pcDNA3 (21a), pCDU4 (21b), pCDU3 (21c) pCDU2 (21d) and pTB72 (21e)

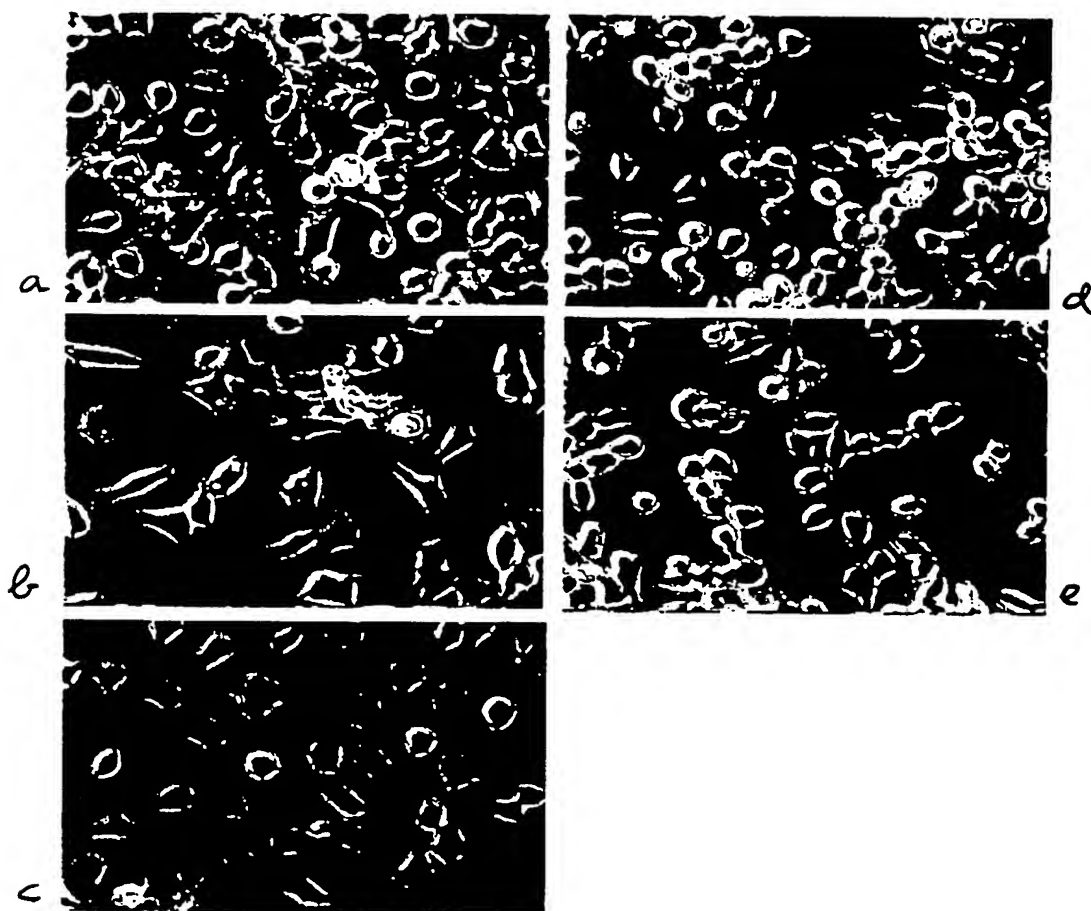


Figure 22 : Phase contrast image of N4 neuroblastoma cells, transfected with pcDNA3 (22a), pCDU4 (22b), pCDU3 (22c) pCDU2 (22d) and pTB72 (22e)

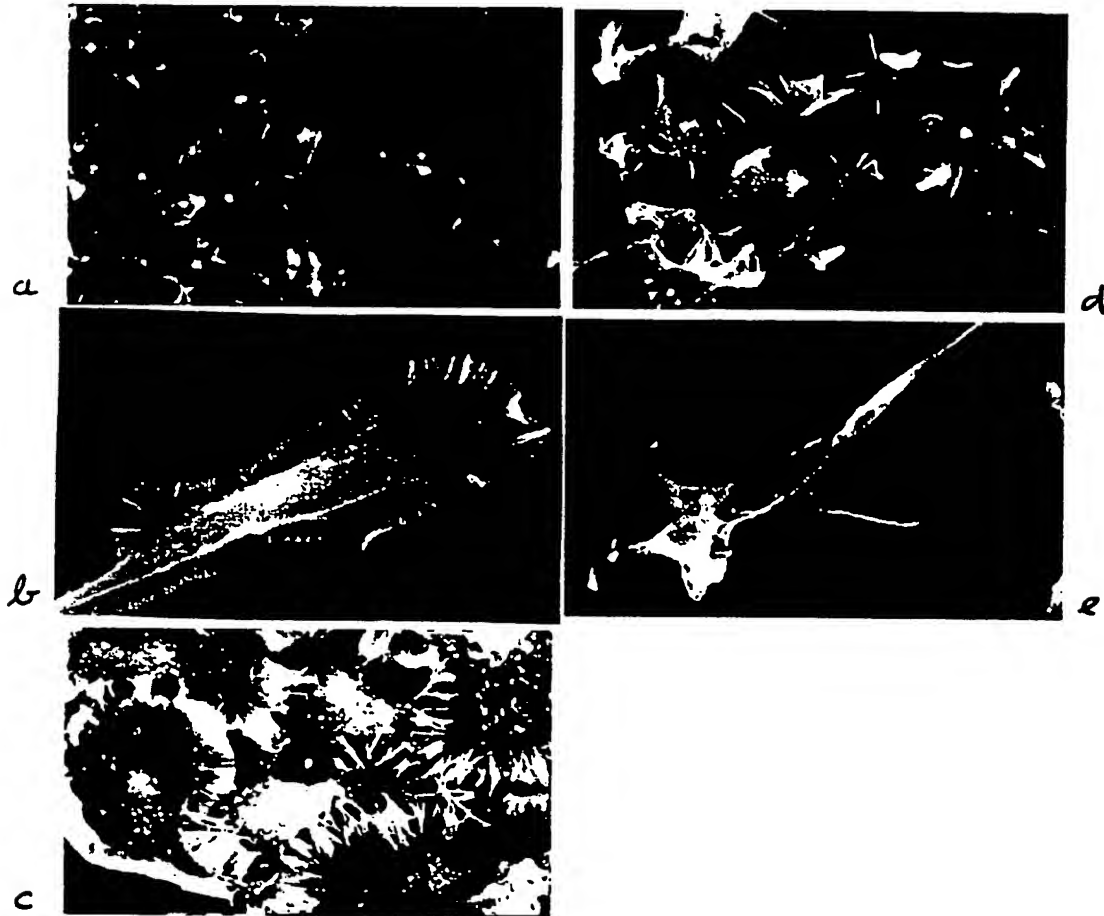


Figure 23 : F- actin pattern Phalloidin (visualized with TRITC-Phalloidin) of N4 neuroblastoma cells transfected with pcDNA3 (23a), pCDU4 (23b), pCDU3 (23c) pCDU2 (23d) and pTB72 (23e)

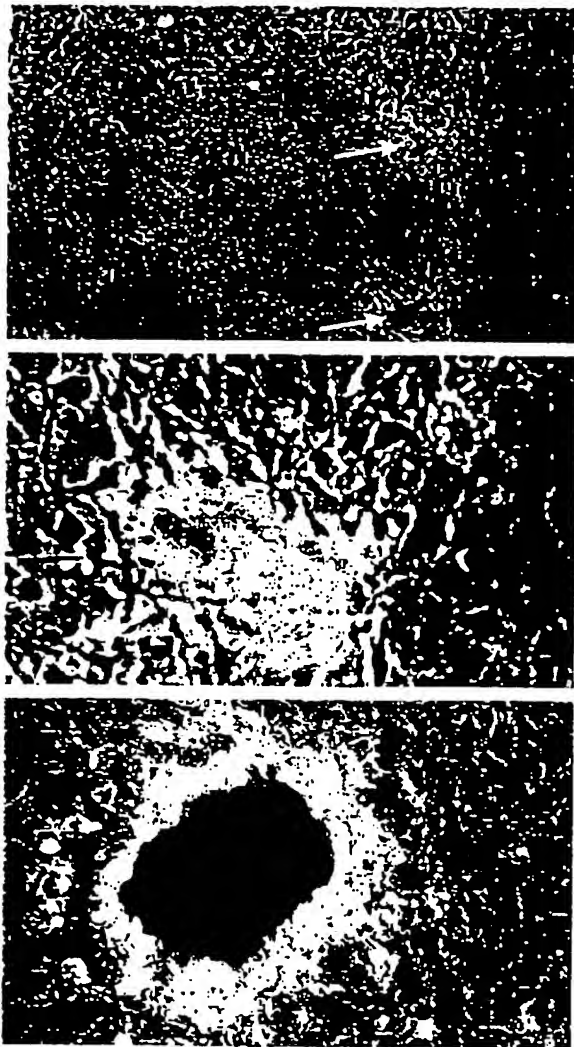


Figure 24 : Phase contrast images of small, medium sized and large foci induced in a monolayer of NIH-3T3 cells by transfection with pCB201.

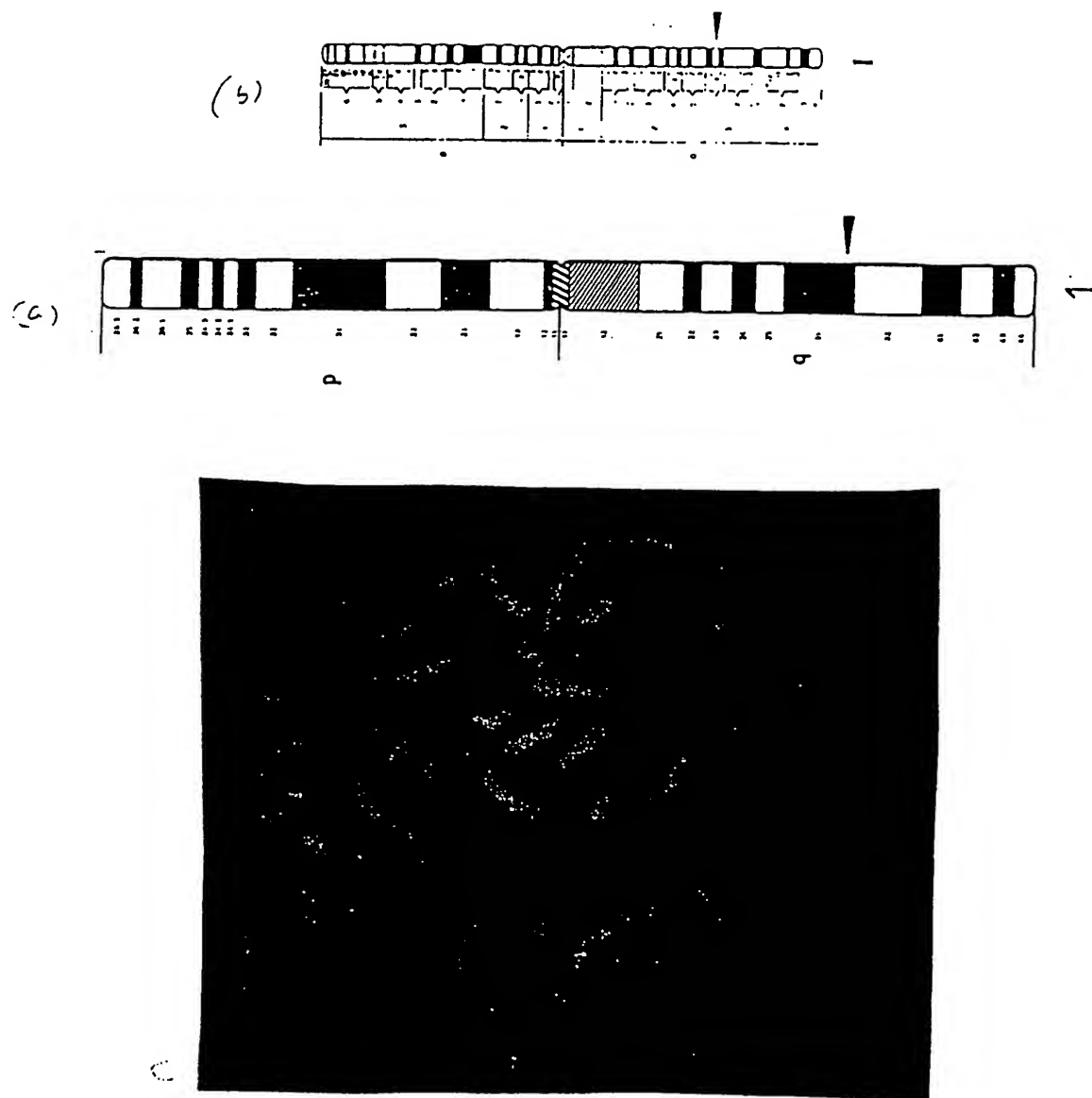


Figure 25a, b, c: Chromosomal localisation of hu-unc-53/1 by FISH

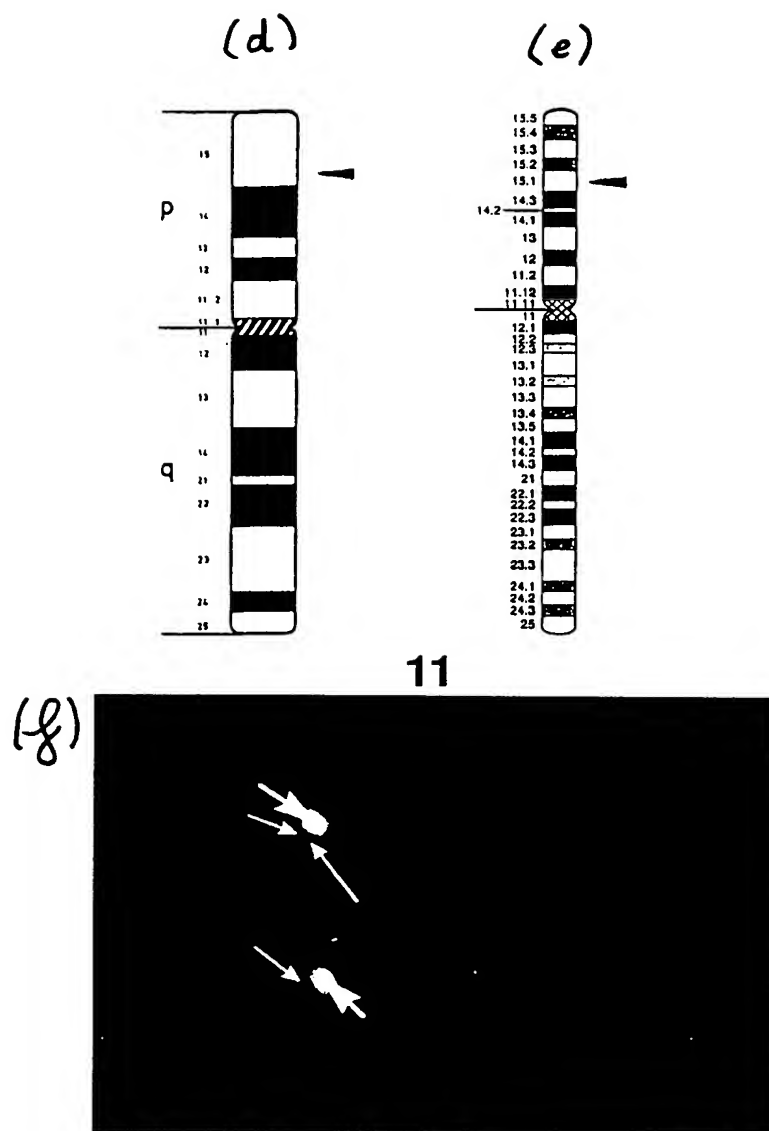


Figure 25d, e, f: Chromosomal localisation of hu-unc-53/2 by FISH

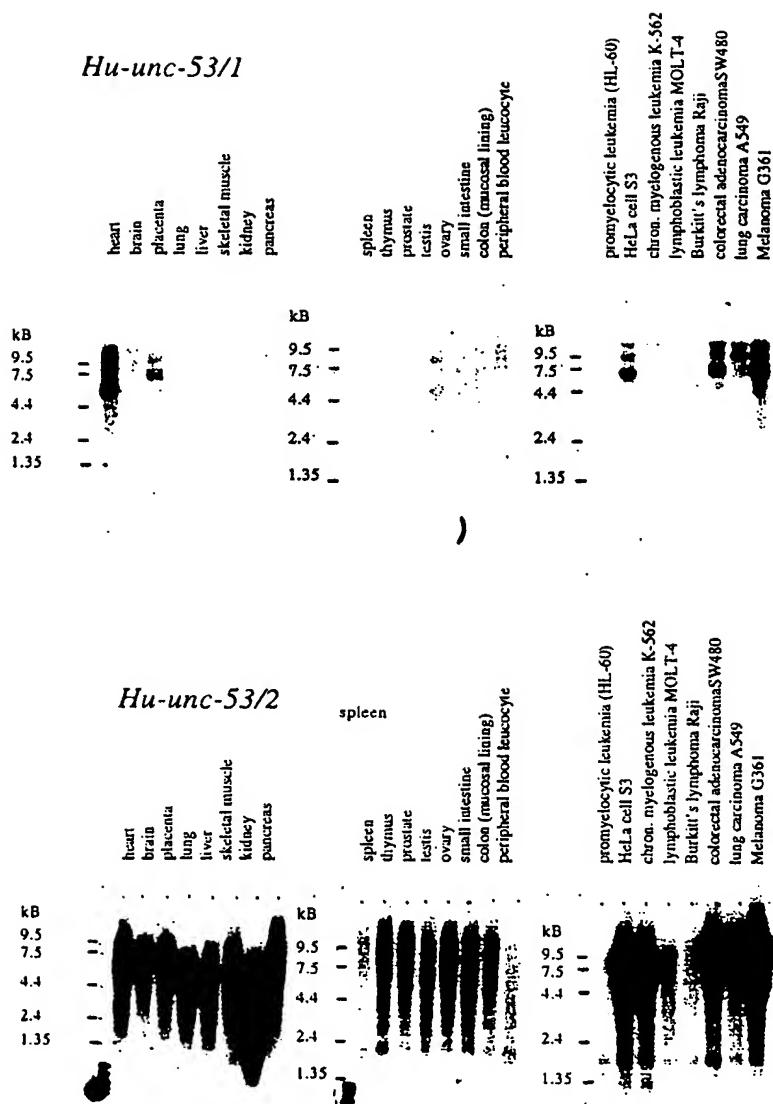


Figure 26 : Expression pattern of *Hu-unc-53/1* and *Hu-unc-53/2* in normal human tissues and cancer cell lines by Northern blotting.

donderdag, 27 november 1997 16:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence
Enzymes: All 146 enzymes (No Filter)
Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

CTAAATGTAGCGTTAATATTTTGTAAATTCGCGTTAAATTTTGTAAATCAGCTCATTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTAT 100
AAATCAAAAGATAGACCGAGATAGGTTGAGTGTGTTCCAGTTTGGAAACAAGCTCACTATTAAGAAGCTGGACTCCAACGCTCAAGGGCGAAAAA 200
CCGTCTATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCAAGTTTGGGGTCGAGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGAG 300
CCCCGATTTAGAGCTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGGAGCGGGCGCTAGCGCGCTGGCAAGGTGATCG 400
GTACGGCTGCGCTAACCACACACCCGCGCGTTAATGCGCGCTACAGGGCGGCTCCATTGCGCATTCAGGCTGCGCAACTGTTGGGAAGGGCGAT 500
CGGTGCGGGCTCTTCGTATTACGCCAGCTGGCGAAAGGGGATGTCTGCAAGGCGATTAGTTGGTAACCCAGGGTTTCCAGTCAAGCGTTG 600
TAAACGACGGCGAGTGGCGCGTAATACGACTACTATAGGGCGAATTGGAGTCCACGCGGTGGCGGCGCTCTAGAACTAGTGCATCCCCGGG 700
CTGCAAGAAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGAGCGGGGACGCGCAGGACACACAGTTTAGTTATTTATGAGGAAGGGGAGAAAG 800
ATTATGCAATGACAAAGAGAGGAATGGTGACAGCCACAGTGTGGCAAGTTCTTCTTGTCAAGTGGCTACTCGAACACATACTATCAATTGGTT 900
TTTGTGTGTGTGTTCCAAATATGGAGGCACACTGGTTGAGGTTGAGGAGCGACTTCAAGACTTATACCTTGGAAACAAATGAAAAATCAAAAAA 1000
AGTCAACTAGGATTTCAATTAATAAATTTCAAACTATTAGATTTTATAAGTTTCTAAATTAGAGTTGCTTCCAGATAAATGGATCAAAAAAT 1100
CTAGGCTTCAATTTTGAATTTCTCGAAGTGATCAACTCAATTTGGCTTAATATTCAAAATTTGAATTTTAAATTAATAAAAAAACCTTTTCTT 1200
TCTTAACTGTTCTAGTAAAAAGCTCCAAATTTTGTGATAATATCTTAGAAGTACATTTTATCAGCTCATCGGGAGCTTTGGCACCTATGG 1300
GCTCTGCTAGGCCCCCACAATAACCTTCGAGGAAGTGAATTTTGTGGCGAGACCCGGAACGTAGTCATGATGGGAGTCAATGGAAGGGCGAT 1400
GGAATTCAAATTTGGCAGTTTAATCTTTAAAAAATTTGGAATTCGAAATAATGATTTGGGAAATTTTGAATAAACGTTCTAAGGTTATAGTTGAGA 1500
TAATTTGAGATCCGAATTGACACACATTTTGGCTAGGTTTGAAGAAGTGGGTCCCGCCACGAAACCTCCCTACAAAGTTTATGAAACAAATTTT 1600
GAAATAAAACCAAGATTTGAACACAAATGCTGAGCAACATGGTGATACTAACATAATTTGGCAAGCTACCAAAACCAAAAAATCTTGACAAACACTT 1700
CACTATCAAAAAAGAAATCAATTCGCGCAAAAAGATGGTAAGAGGAGAAAAAGAAACCCGCTCTGTTGTTCTCGGCTTAATGAATTAATTATAGGAAC 1800
ACAAGGCAATCGGAAGGAAACGGCGACGGAAGGGAAGGAAGGAAGCAAGTCCCAAAATGGAAGCGGGAGGACAGCTGCATTAGAGAGAGAGGGGAG 1900
CTTGTGCGCCAAATCTAGGGTGAATGAAAGAGATTGAAACCAACAGATCGGGTGACATGAAAAGAAATGAGACGAAGAAGATGATAGGTTCCG 2000
GTTATTTCAATCCTAAATTAATGGACAGGACTTGAAGGAAATTTGAATTTAGGCAAGGCTCAAGCGGAACCTTTTGAAGCTATAGTATGCAGACTC 2100
ACAGAACTACAGAAATTTCAAAATTCATGAACTTTCTGAATACCAACTATCTTGCCCATCTCCACTGAAACATCTCAAAATTCATAACTAGTTTCAAAT 2200
CAGAGGCATAACAAAAAGCAACAAAAAGCAAAATGCTCTCGAAGCGCGCAAAACCGCATCGGTTCCATTATGAGATGCACCGATAGATCTTCGTTA 2300
GAAAGAGAGAGAGTAAATAACATTTGGATATCCTCTAGGAGACAACCTGCGCTAAACAGATGTTGGTTCTAGTATTTTGAATAATGCAATTTGGCG 2400
GCAACTTCTTCCCAAGTTTGGCCCAAAAAGTATTGACCAAGCTTGGTACAGACTCTAGAAAGACTTGGGCGAGCAAGACTTGTGAGCAAAATTTTGC 2500
CACTGTTTGAATTAACGCTGACCCAAACAAATTTGGCTCAAGTTTACCAAACTGCGGCAAAATGCAAGTAGACTGAGGTCAACCGAATATCTCTG 2600
TAATTTGCAATCTAAGCTCCAAACTATTTCAAGCGGTTCTGATGCTCATTGATAGTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700
GCACTTCGCGCATGAATAGATGTCTGTGAACAGGAGCTCCAGCAGAGAGAGAAAAAGATGCAAAAGAGAGATGCTATCAAAAGTCTTTCGGGGAT 2800
TCCTCGATACCAAGTATGCTCATTAATTTGAGCTTCTGGATTAATTGCTCCATGGCTCCCTGACCAACCAACGAGCAGCGAATGTGTTCATATG 2900
GTGTCAAGGAGACTACAACACTACTACCAAAACTCACTGTGAGGATAGCGGTTCTGTGCTCGACGAGTCAATGCTATCTTCTCTCTCTCTCTCTCT 3000

donderdag, 27 november 1997 16:52
fig 27 pNP3 Map (1 - 13621) Size and Sequence

Page 2

ATTTTGGCCACATCGCAACACCGCGTCAGAGGTGTAGCCGCTGGATTCTGAAATATTTTGAGAGTTATGAGGATTCAACTGTGGAAGGTTTTTTTTCA 3100
AATTCTAACGCCAATGTGCTCAAAATTAATCTGTGAAATTAATAATCTACAGTGTTCATTCACCGAATTAGTGATTAACGTTCAAAATGTAGACAACAA 3200
GGAATAGGTAAGCATGTAGGCAATGTAAGGTGTTGGTAGGTCCAATAACAATAATTATTAGCTTTTGATTATTCAGTGTTTTTTTTAAATATTATT 3300
TACTGTTTATTGGCTTGTAAATATCAAGCCTATCTGCCTACATCATCTTTCTGCCTACACACCTGCCTATGCCCTAAACTTAATACATCTCAAGAGA 3400
ACTTTCACAAATTCAAAAATTAATACTATAGATTCAAAATTAATCAAAAGAAAAAGTTTTTGGTACGAGCTCCAACGTTGGTTAACTTTAAACGAAAC 3500
AGTTCAGAAACCAAAAAATCAAAAGTTTTCTTTGACAAACATCATTTTTCTGGGATTTTTTAAATTTATAATTTGAAGTATTITTCACAAAAAAA 3600
TCAAGTTTCAAAATTTACCTGATGTCTTATCTTTCCCATCTTCTCTCCAGCGAATGATTACCAGAGTTTTTGTGCGCGGAGTCACCTGAGCCGAC 3700
GCATGAGCCCAATGGAGTCTTCAGACGCGGAGGACTGCTCATATCTGCAAAATGGTCATTTGTCTAATAATTTGGAGGTGGGGGTGCACTCACCAAA 3800
CGACTTGACAGGCGATGTGATTTTTCTTGTTTTAGAAGAACATCGTACGATAAGGTATTTTGTGCGGTTTTTCTATGATGTTTTCAAAAAAATTTA 3900
TCAGAAATAAAAAACAATAAGAGTGTACACAGAACATTTGTGTAGATTTTGGGATAGAAATTTATTAATGGGGGAGGGAGCAATTTTTTTTCACT 4000
TTTTATTACGAATTAATTTCTATTTTGAATCTACACAATTTCAATTTAGAGGTGACTTAAAGAACTAAAAAACAAAGAAATTAAGAAAGATTTTTT 4100
TGATAATAACAAAAATAGGGGTGGGAGCAGGAGGAGAGCATGTAGGAGTGTGAAGAGTAATTGGACTTAGAAGTCAGAGGCACGGGCGCGAGAT 4200
GCTGAAAAATTAATATTACTCGTTAGTTAGTAGACTCAGTTTAACTTACTCTGATGACAGCGGCCCTATTATTTTGCACACAGACAAGTTGGTAA 4300
TGGTAGCGACCGCGCTCAGTTGGAAATCTACGAATGCTATTTGTATAGTTTCCATGCCATGTGTAATCCAGCAGCTGTTACAAACTCAAGAAAGAC 4400
CATGTGGTCTCTTTTTCTGTTGGATCTTTGAAAAGGCGAGATTGTGTGGACAGGTAATGGTTGTCTGGTAAAGGACAGGGCCATCGCCAAATGGAGTA 4500
TTTTGTTGATAATGGTCTGCTAGTTGAACGCTTCATCTTCAATGTTGTCTAAATTTGAAGTCTGAAAAATTAATATAATCTCGTTAGTTAGTAAGT 4600
CCAAGTTTAACTTACAACCTTGATTCCATTTCTTTTGTCTGCGCATGATGTATACATTGTGTGAGTTATAGTTGATTTCCAAATTTGTGTCCAAGAA 4700
GTTTCCATCTCTTTTAAATCAATACCTTTTAACTCGATTCTATTAACAAGGGTATCACCTTCAAACTTGACTTCAGCACCTGAAAAATTAATATGTAT 4800
GGTAGTTAGTACCGAACTGTTTAACTTACGTGTCTTGTAGTTCCCGTCATCTTTGAAAAATATAGTTCTTTCTGTACATAACCTTCGGGCATGGCAC 4900
TCTTAAAAAGTCATGCCGTTTATATGATCTGGGTATCTCGAAGGACATTGAACACCATAACAGAAAGTAGTGACAAAGTGTGGCTGAAAAATTTAATA 5000
ATCAGGGTTAGTTAGTATATATATGTTTAACTTACCATGGAACAGGTAGTTTTCCAGTAGTGCAAAATAAATTAAGGGTAAGTTTCCGTATGTTGCA 5100
TCACCTTCACTCTTCACTGACAGAAAATTTGTGCCATTAAACATCACCATCTAATTCAACAAGAAATTTGGGACAACTCCAGTGAAAAAGTTCTTCTCTT 5200
TACTCATTTTTTCTACCGGTACCTCCAAGCGTACGTTGGTGGTGTGCGATTGAAGTCATTTAAGTGGAACTGATTTTAGAGGTGGAAAGCGGTGG 5300
AACGTCGTCGTGTCGATGCTCATTGATTGGAGCTTTTCAAGTTCTGGTTTTTCAGATATCACGGGTCTGAAATTAATAATATGAATTTGAGATTGAG 5400
GAGCATCTCTGAAAAATACAAATATTTTAAGATTCTAGTTTGACAAAAATTTATAGCCCAAGCTGAAAAATCGCCGAAAGTCTGGAAAAATGCTTGGT 5500
TTCAAGAAATCACATTTTAAATTTTTTAAATTTTCTCAAGGAAATTTTTTCAATTTCTGGAATTTTCTCAATTTGTTTTTACAATTTTGGAAAT 5600
TTTTCTCAATTTTTTTCGGAATAAGTGAATTTTTGATAGTTTTTCACTATGCTCTATTTCTGAAAAATGAATCTTTTGATATTTTTTTTGTGAT 5700
TTTTTGAAGTAGCTTTATTTGTGAGTTTTTCAAAATTTTTTTTTGGAATTAATTTCCGTTTTTCTGATTTTTTGTGCTGCGCAACGAAAAATC 5800
GAATAAATTAATATGATTTTAAATATGTTTGTAAATTCACAAACAAATCTTACTATTAGACATATTTGGAAAAACATTTTGAAGAAATTTCCCAAG 5900
AATTTAGAAATTTTATAGAGTATTTTATTTTTTAACTCAATTTTAAACAACTTCAGAAAAAACTTATGGTCAGAAAAATCAAGTTTTTACT 6000

donderdag, 27 november 1997 16:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence

Page 3

GAACCTTCTACAAATAAAATAATTGTACTCTCTTACTATYCCCGCACAACTTAAATAATCCAAAAATTACTCCCATCATGTGTCAACAAAAACG 6100
TCACCGACCGAAACATAAGTAAAGTTCAGAAATAAACGTACCATTTGTCACTCTTATGTGCATAATTGGACTAACCACTCGATTGTGGCTGGGTG 6200
TCACGTGGCGGAACAGCTGGAGGTGATCTTTTTCCGTGTGCTTTTCACTCTTTCACTGCCAGTGTGGTTTTCTAAAAATTGATTATTTTCGAA 6300
AATATAGGAAGCTCTTAGGAAGCCGAAGCCAGACATTAAAGTGTATCAAAATCTTAAATCTAAAAAGGGAAAAATCTGCAACAAAAGACGCAAAAAA 6400
GGAGAAATGGAAGACGGAAAGCGTATTGAATTCAGATATTTCCATTTTGCACCTTTGCTAGTCCGACTTTGAAGCCAGCTTGAATCTTTAGAT 6500
TGCACCTACTCGGCGCAAGAATATGTACATGGTTCAGGCCGGCCAGTTTAGCCCATGCTTGATCAATGGCTTTGCCAGCCTCTAAACCAAGTTAATAT 6600
AATTTAGGCCAAACTTGACAAGAAGCTCTACCCAGAAACAAAAATACTCACCTCAACAGGCTTGTGATAATATTTGGAGAAACCGGTGTTCGGGCTA 6700
TCGGCTGTCTGATAGCTGTACGATGGAGCGTTAAGAGTAAGATCGTCTGATGACCGAGACTTTTCGTCTGACGTTGAACCTTGAAAAATTTGAAAGTTT 6800
TATAGAAAGGGGACGGTTTCAATGGGATAAATTGAACTTAAAGTTTCTCTCTAAAAATTCAGTTGCACAAAAAATTAGTTTCAACTTATAAAATGAG 6900
ATTATTAATTCGAAAGCTTCTCGCGCACCAAAAGTTTATGGGACGGAACTCTATCTGAATGTTCTAAAAATTTTATCCATATAAAAAAGTTTAAAGTA 7000
AAACATTTTTCTGTAATATCTTAACTTTTACATCAATTTTGTCTGAAAAATAAAATTTAAAAAGTTTGGGTTTGACAAATACCGTATGACATTTAT 7100
CTTAAAAATGTATAAAGCAGGTCTCTGCAAGTTTTTTGTACCAACAAATGGCGGATTCACATTTTATGGCAGGTGTCCAGAGACCTCACTTTTATA 7200
AACATCCGAGACTATGTGAAGTCAAGAAACAAAAAACAAGTTCTCTTAACAACGAACCTTGAAGATGTGAATGCATGCTTAGGGAACCTTCCG 7300
TCGATGATGACGTTGGCGAGCTGTGTGAATCCAGCGTATCCGGACTCTTCTCACTGCTCTTTGAGCAGCTGTTTCGAGCTTGTTCAGATATGATTGG 7400
GGCGTCCGTTTTACGGTAGGAACTTTTGGCTGGAAATTACTATGTAAAAACGAAGTGCACAGTTTTTTTTTCAAAAAATTTTTAAAAATGTTTTT 7500
TCATAAATAATAGAACCCATAAGACAATTCCGGAAAGTCAAAAGAACTCTGTATATTTCACTAGGAAGTTTTAAAAAATATAGCAGGTCTAAGAAAGCTCT 7600
AATTTATATGAAAAAATAAATACTAGAAAACGAAAAGTAAAGTTAATTTTTCAAAAAGTGCATCTCACAAAAATGTTTTGCAATTTGTGCTCTTTC 7700
ACTCTGTAGTAGCTCATTTTGAAAAACGAATTGAAAAACCGGCGAAGAAGCTGAGAAATGCTTTACTCTTTGTCAGCCAGATGTGTTTTTCGGCCCT 7800
CGACCTGCAAAAAATGCTCGGAATATCACTTCCGTTGCAACATTTTCCGCGAAGAGCTTTATGGGGGGACACAACGACAAAAAGATGGGACATAGATTC 7900
TGAAGAGTGTCAAGGAGAAAGAAAGAAATGGGAGACAAATCCCAATTGTCTCATCTTTTATAGTAGGTTGTTTTGTTTTGGTAGAGTCAAGGAA 8000
CGGAAGAAGTACGAATCGGGTTCTAAATATAGATGAGGCCAGAGCTTCGCAAACTAGGGTTTTAAATTTCTTGAAATCAACTTTGACAGCTTATAT 8100
CTCACTGTCTTTGGTCTATACAAAAATATCAACTGTCAAAATATTTTTATTTCTTTTCTATATCTGATATCTCAATATGAACCAAGATATAAAT 8200
CTTCAAGTAAATAGTGGGATGCAATCTATCAGAGGAAAAATCTATTTTAAATTA!TTTGTAGAAATTTCACTTTTAAAAAATACATTTCCAAT 8300
TTCCGATTTCAAAATCTCACCGTACAAAGCTTCGACATAGACGTGGCACTTCCAGCTTACTGGTCCGGCGCTTCAGGCCACTTTTCACTGGGGCAGCGA 8400
TTTTGACAAAGTTTSTTGTGAGGCACCGCCGCCCTTTCTGTAGGTGTGGGCTATTCGATGAGGAAGATGGGTTTTTGTACTGAAATAACTTTAA 8500
TTTCAGCATTCACCACCACCACCACCGCTGTATCGGCTCTTGTTTTGTCTCAATAGTCTTCACAGAAGCAAGTTTGGGGTGTCAAGGCTTTCCGA 8600
GGCGTAGCTTTGAGCTTCGATTTTGTAGTTGTAGCAACCAACTAGCTGGGTTTGTGGTCTAGAAGGTTTTTGGAGTTGGGAGGTAGGTGGTTTA 8700
GATTGCAATAGAGCTGTACGTTGATGATGATTCTGAAATAATAAATTTCTAAAAAGAGCCGGAAGGATCGGATACCTAAGCTCTTCGCGAGATGTGGA 8800
TATCTCGAGCCAACATTATTAATTCCTCGAACGGCTCGACGACGGAATGAATTTGTATTATTTGATGAAGTGGTGTATGATGAGGCTGGTTTAAGT 8900
CCAGAGCTCTTTGGCTTGATACCAATCTTTGATGAATCTGTAATTGAAAAGTGTGAGAGAAAGGCATAACGATTTTACGCATCACAGGCAATTAAGATTTT 9000

Page 4

[illegible]

dondomg, 27 november 1997 16:52
fig 27 pNP3 Map (1 > 13621) Site and Sequence

Page 5

GTTCGACCTGCCCTTACCGGATACCTGTCGGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGT 12100
AGGTCGTTCCCTCAAGCTGGGCTGTGTGCAGAACCCCCGTTACGCCCCGACCGCTTATCCGGTAATATGCTCTGAGTCCAAACCCGGTAAG 12200
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTTCTTGAAGTGGTGGCCTAACTAC 12300
GGCTACACTAGAAAGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAGTGGTAGCTCTTGATCCGGCAAAACAAACACCG 12400
CTGGTAGCGGTGTTTTTTTGTTCGAAGCAGCAGATTACGCCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTCTACGGGGTCTGACGCTCA 12500
TGGCAACGAAAACTCAGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTAAATTAATAATGAAGTTTAAATCAATC 12600
TAAAGTATATAGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGTTTCATCCATAGTTGCCT 12700
GACTCCCCGTCGTGTAGATACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC GGCTCCAGATTT 12800
ATCAGCAATAAACAGCCAGCCGGAGGGCCGAGCGCAGAAAGTGGTCTGCACTTTATCCGCCCTCCATCCAGTCTATTAATTGTTCCCGGGAAGCTAGA 12900
GTAAAGTAGTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTGGTTGGTATGGCTTCATTCAGTCCG 13000
GTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCTCTCGGTCTCTCGATGTTGTGCAAGTAAGTTGGCCGC 13100
AGTGTATCACTCATGTTATGGCAGCACTGCATAATTCTTTACTGTCTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCA 13200
TTCTGAGAAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGCTCATCTTG 13300
GAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAAACCACTCGTGACCCCACTGATCTTCAGCATCTTT 13400
TACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCCAAAAAAGGAATAAGGGGACACGGAAATGTTGAATACTCATACTCTTC 13500
CTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAAACAAATAGGGGTTCCCGCA 13600
CATTTCCCCGAAAAGTCCAC 13621

FIGURE 28.

SIGNATURE SEQUENCES :

Different signatures can be used to define to identify the UNC-53 gene family :

Aminoacids are listed in one letter code

X equals any aminoacid

X(3,5) equals 3 to 5 X's

(D,E) means D or E at a given position

The signatures should be used to screen a database using a weight matrix of conservative substitutions.

BLOCK A :

GSXLSLASX(3,5)YXSXXEX(4,5)IRXXXR(D,E)LEXXXVXXLTXXXXXXXXLX
XXFEQXL

BLOCK B :

KKKKSWXXXXXXXXFXK

BLOCK C :

LARGE FAMILY :

VXXL(K,R)XELX(D,E)(R,K)(D,E)XXLXX(V,I)RL(D,E)XLXXAXXXDXLRE(T,A)X
XXXXXEXXXLKXEXD(R,K)LX

VERTEBRATE FAMILY :

VXXLRXELX(D,E)(R,K)(D,E)MKLTDIRLEALXSAHQLDQLLREXMXNMQXEXX
LKAENDRLK

BLOCK D :

LARGE FAMILY :

W(K,D)X(I,L)DXX(I,V)XX(L,V)F(K,E)XY(I,V,L)XXXDXXXLG(I,L)X(2,3)(D,E)S
(I,V)XGYXI(G,S)(E,H)(L,V,I)(R,K)(R,K)

VERTEBRATE FAMILY :

FXXGCXXVSGKXXWXXLDXXVXX(L,V)FK(D,E)YIXXXDPXXXLG(I,L)XX(D,E)
S(I,V)XGYSI(G,S)XXKR

BLOCK E :

GXXGXGKS/T

and

F(K,R)MXXXSXX(3,8)GF(I,L,V)(I,L,V)(R,K)Y(I,L,V)(R,K)(R,K)(R,K)XV(D,E)

and

F(I,L)EKXXXX(D,E)XXXGPPXX(L,I)XCPXXXXXX(R,K)XWFXLWNXXIPY(
L,I)XXXA(R,K)DGX(K,R)XXGXXXX(F,W)EDP

Block F

(W/F) (D/E) DSSS (V/L/I) SSGISD (T/N)

Tuesday, 18 November 1997 10:34

fig 29 pEGFPsac (1 > 5100) Site and Sequence

Enzymes : 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

7p.

Bgl I

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGGTTACATAACTTACGGTAAATGGCCCCCTGGCTGACCG
ATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGTATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGC
L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

Aat II

CCCAACGACCCCCCGCCCATTGACGTC AATATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGT
GGGTTGCTGGGGGCGGGTAATGTCAGTTATTACTGCATACAAGGGTATCATTCGGGTATCCCTGAAAGGTAAGTGCAGTTACCCACCTCATAAATGCCA
A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

Aat II

AAACTGCCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTC AATGACGGTAAATGGCCCCCTGGCATTATGCCAGTA
TTTGACGGGTGAACCGTCATGTAGTTACATAGTATACGGTTCATGCGGGGATAACTGCAGTTACTGCCATTACCGGGCGGACCGTAATACGGGTCA
N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

Bgl I

CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTGGCAGTACATCAATGGGCGTGGA
GTACTGGAATACCTGAAAGGATGAACCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACATCGCCAAAACCGTCATGTAGTTACCGGCACCT
H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V

SnaB I

Aat II

TAGCGGTTTGACTCACGGGGATTTC AAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ATCGCCAAACTGAGTGCCCTAAAGGTTTCAAGGTTGAGGTAAGTGCAGTTACCTCAAAACAAAACCGTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCAT
I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .

Nco I

ACAACTCCGCCCCATTGACGCA AATGGGCGGTAGGCGTG TACGGTGGGAGGTC TATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
TGTGAGGGCGGGTAAC TGCGTTTACCCGCCATCCGCACATGCCACCCTCCAGATATATTCGCTCGACCAAAATCACTTGGCAGTCTAGGCGATCGCGAT
O L R P I D A N G R . A C T V G G L Y K O S V F S E P S D P L A L

Nhe I

Nco I

CCGGTGGCCACCATGGTGAGCAAGGCGGAGGAGCTGTTTACCGGGGTGGTGCCCATCTGGTTCGAGCTGGACGGGACGTAACGGGCCACAAGTTCAGCG
GGCCAGCGGTGGTACCACCTCGTTCCGCTCCTCACAAGTGGCCCCACACGGGTAGGACCAGCTCGACCTGCGCGTGCATTTCGCGGTGTTCAAGTCGG
P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

Nhe I

ATCGGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCAACCGGCAAGCTGCCCGTGCCCTGGCCCCACCTCGTGAC
ACAGGCGCGTCCCGCTCCCGCTACGTTGGATGCGGTTTCGACTGGGACTTCAAGTAGACGTGGTGGCGGTTCGACGGGACGGGACGGGTGGGAGCACTG
V S G E G E G D A T Y S K L T L K F I C T T G K L P V P V P T L V T

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 2

CACCC TGACCTACGGCGTGCA GTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTCTGCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATGCAGGTCTCT
900
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E
KspI
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTGTATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCAC TTGGCGTAGC TCGACTTCCCGTAGC
1000
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I
ACTTCAAGGAGGACGGCAACATCTGGGGCACAAGCTGGAGTACAACACAAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
TGAAGTTCTCTGCGGTTGTAGGACCCGTTGTCGACCTCATGTTGATGTGTGCGGTGTTGCAGATATAGTACCGGCTGTTCTGCTCTCTTGGCGTAGTT
1100
D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I
GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCATCGGGACGGCCCCGTGCTGCTG
CCACTTGAAGTTCTAGGCGGTGTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCGCTTGTGGGGGTAGCCGCTGCCGGGGCACGACGAC
1200
V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAGACCCCAACGAGAAGCGCATCACATGGTCC TGCTGGAGTTCTGTGACCSCCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTTACAGGACGACCTCAAGCAC TGGCGGGCGCCCT
1300
P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G
BspM II Bgl II
TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTCAAGCC TGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCC TAACCCGTTAGCCGTGGAAG
1400
I T L G M D E L Y K S G L R S T S N V E L I P I Y T D V A N R H L S
C.e.unc53 sac
Nru I EcoR I
GAAGGSCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGCGACTATCGACTGGTTTTCAGCTTATTAATGTATCGTTCGGATCAACGAA
CTTCCCGTCGAATAGTTTACGTAATCCCTATAAAGGTTACTAAAAGCGCTATAGTGCACCAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT
1500
K G S L S K S I R D I S N D F R D Y R L V S Q L I N V I V P I N E

Page 3

Bsm I

TTC TCGCCTGCATTACGAAACGTTTGGCAAAAATCACATCGAACCTGGATGGCCTCGAAACGTGCTTCGACTACCTGAAAAATCTGGGTCTCGACTGCT
AAGAGCGGACGTAAGTGCTTTGCAAACCGTTTTTAGTG TAGCTTGACCTACCGGAGCTTTTGACAGAGCTGATGGACTTTTTAGACCCAGAGCTGACGA
.....
.....
.....

C.e.unc53 sac

F S P A F T K R L A K I T S N L D G L E T C L D Y L K N L G L O C

EcoR V **Pvu II** **Ear I** **Ksp632I** **Hind III**

CGAAACTCACCAAAACCGATATCGACAGCGGAAACCTGGGTGCAGTTCCTCAGCTGCTCTTCCTGCTCTCCACCTACAAGCAGAAGCTTCGGAACCTGAA
GCTTTGAGTGGTTTTGGCTATAGCTGTGCGCTTTGAACCACGTCAGAGGTCGACGAGAAGGACGAGGTTGGATGTTTCGCTTCGAAGCCGTTGACTT
.....
.....
.....

C.e.unc53 sac

S K L T K T D I D S G N L G A V L Q L L F L L S T Y K Q K L R Q L I

Sst II **Xma I** **Apa I** **Sma I** **Bam HI** **Xba I** **Bcl I**

AAAAGATCAGAAGAAATTGGAGCAACTACCCACATCCATTATGCCACCCGCGGGCCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATACC
TTTTCTAGTCTTCTTTAACTCGTTGATGGGTGTAAGTAATACGGTGGCGGCCCGGGCCCTAGGTGGCC TAGATCTATTGACTAGTATTAGTCGGTATGG
.....
.....
.....

C.e.unc53 sac

K D Q K K L E Q L P T S I M P P A G P G S T G S R . L I I I S H T

Dra I **Bsm I** **Hpa I**

ACATTTGTAGAGGTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTS
TGTAACATCTCCAAAATGAACGAAATTTTTGGAGGGTGTGGAGGGGGAC TTGGACTTTGTATTTTACTTACGTTAACAACAACAAATTGAACAAATAAC
T F V E V L L A L K N L P H L P L N L K H K M N A I V V V N L F I

Bsm I

CAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTCTACTGCATTC TAGTTGTGGTTTGCCAAACTCATCAA
GTGGAATATTACCAATGTTTATTTCTGTTATCGTAGTGTTTAAAGTGTTTATTTCTGTA AAAAAGTGACGTAAGATCAACCAACAACAGGTTTGAGTAGTT
A A Y N G Y K . S N S I T N F T N K A F F S L H S S C G L S K L I H

Mlu I **Ssp I**

TGTATCTTAACCGGTAAATGTAAGCGTTAATATTTTGT TAAATTCSCGTTAAATTTTGT TAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGS
ACATAGAATTGCGCATTTAACATTCGCAATTATAAAACAATTTTAAAGCGCAATTTAAAAACAATTTAGTCGAGTAAAAAATGGTTATCCGGCTTTAGCC
V S . R V N C K R . Y F V K I R V K F L L H Q L I F . P I G R N F

Tuesday, 18 November 1997 10:34
lig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 4

Bsr I

CAAAATCCCTTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTGTCAGTTTGGAAACAAGAGTCCACTATTAAAGAAGCTGGACTCCAACGTC
GTTTTAGGGAATATTTAGTTTTCTTATCTGGCTCTATCCCACTCACAACAAGGTCAAACCTGTTCTCAGGTGATAATTTCTTGACCTGAGGTTGCAG 3200
Q N P L . I K R I D R D R V E C C S S L E Q E S T I K E R G L Q R

Dra III

AAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGA
TTTCCCGCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACCTGGTAGTGGGATTAGTTCAAAAAACCCAGCTCCACGGCATTTCTGTGATTTAGCCT 3300
Q R A K N R L S G R V P T T . T I T L I K F F G V E V P . S T K S E

Nae I

ACCCTAAAGGAGCCCCGATTTAGAGCTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCT
TGGGATTTCCCTCGGGGGCTAAATCTCGAACTGCCCTTTCGGCCGCTTGACCCGCTCTTTCTCTCCCTCTTTTCGCTTTCCTCGCCCGGATCCCGCGA 3400
P . R E P P I . S L T G K A G E R G E K G R E E S E R S G R . G A

Ksp I

GGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCCGCCGCTTAATGCGCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGAAATGTGCGC
CCGTTACATCGCCAGTGGACGCGCATTTGGTGGTGTGGGCGGCGGAATTACGCGCGATGTCCCGCGCAGTCCACCGTGAAAAGCCCTTTACACGGC 3500
G K C S G H A A R N H H T R R A . C A A T G R V R V H F S G K C A

BspH I

Ssp I

Ear I

Ksp632I

GGAAACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
CCTTGGGGATAAACAAATAAAAGATTTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTACGAAGTTATTATAACTTTTCTTCT 3600
R N P Y L F I F L N T F K Y V S A H E T I T L I N A S I I L K K E E

OxaN I

Pvu II

Sph I

Ava III

Nsi I

GTCTGAGGCGGAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAGTCCCGAGGCTCCCGAGCAGGCAGAAGTATGCAAGCATGCATCTC
CAGGACTCCGCTTCTTGGTCGACACCTTACACACAGTCAATCCACACCTTTCAGGGTCCGAGGGGTCGTCCGCTTCATACGTTTCGTACGTAGAG 3700
S . G G K N Q L V N V C Q L G C G K S P G S P A G R S M Q S M H L

Sph I

Ava III

Nsi I

AATTAGTCAGCAACCAAGGTGTGGAAGTCCCGAGGCTCCCGAGCAGGCAGAAGTATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGGCC
TTAATCAGTCGTTGGTCCACACCTTTCAGGGTCCGAGGGGTCGTCCGCTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGCGCGG 3800
N . S A T R C G K S P G S P A G R S M Q S M H L N . S A T I V P P

Bsr I

Nco I

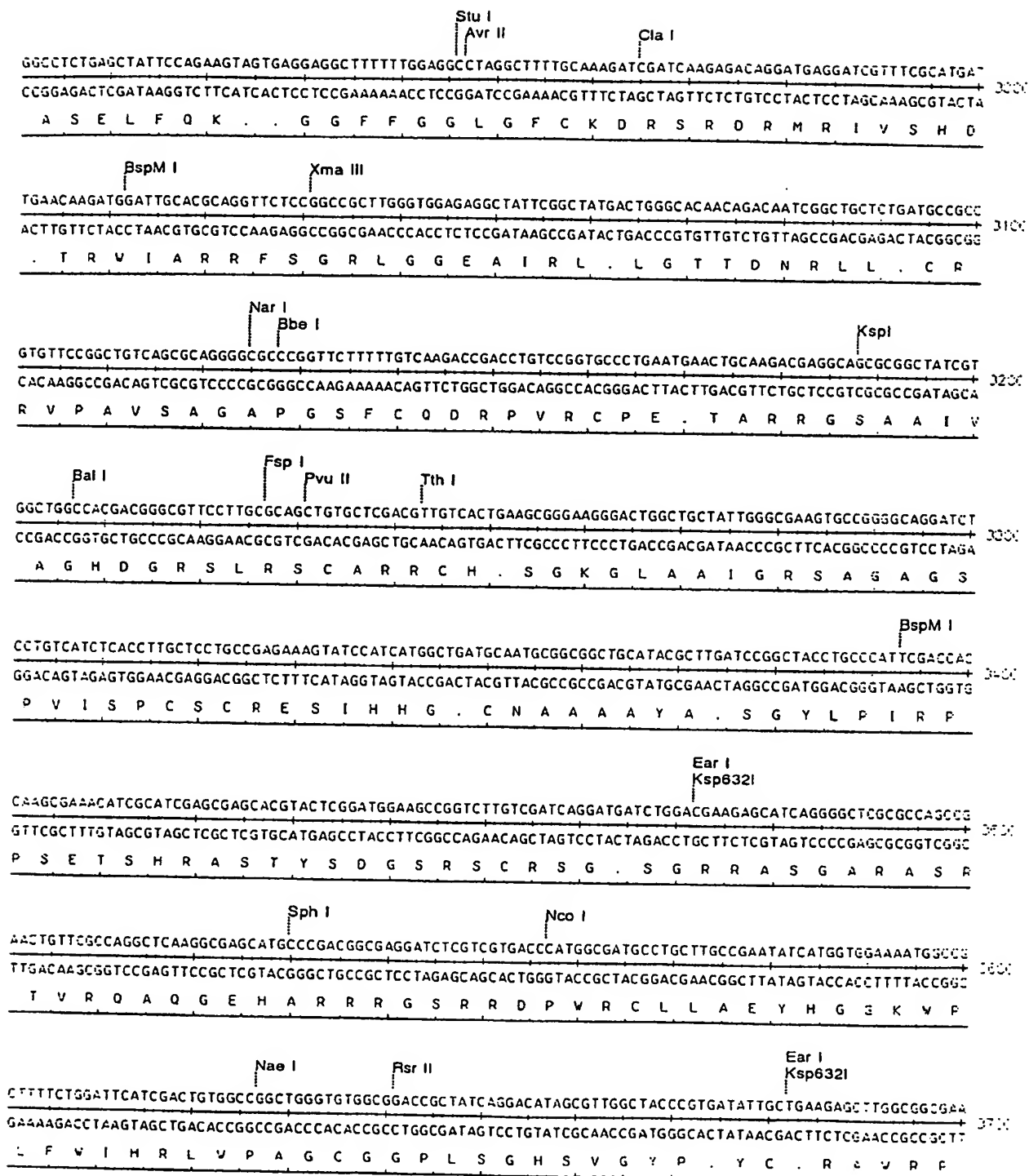
Bgl I

Sfi I

TAATCCGCCCATCCCGCCCTAACTCCGCCAGTTCGCCCATTCGCCCCCATGCTGACTAATTTTTTTTATTTATGCAGAGGCGGAGGCGCGCTC
ATTGAGGCGGGTAGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAGAGGCGGGTACCGACTGATTAATAAAAAATAAATACGTCCTCGGCTCCGCGGAG 3900
L T P P I P P L T P P S S A H S P P H G . L I F F I Y A E A E A A .

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 5



Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 6

TGGGCTGACCGCTTCCTCGTGCTTTACGGTATGCGCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
ACCCGACTGGCGAAGGAGCACGAAATGCCATAGCGGCGAGGGCTAAGCGTCCGCTAGCGGAAGATAGCGGAAGAACTGCTCAAGAAGACTCGCCCTGAGA 380
M G . P L P R A L R Y R R S R F A A H R L L S P S . R V L L S G T L

Asu II BspM I
GGGGTTCGAAATGACCGACCAAGCGACGCCAACCTGCCATCAGAGATTTCGATTCCACCGCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTTC
CCCCAAGCTTTACTGGCTGGTTCGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAG 390
G F E M T D Q A T P N L P S R D F D S T A A F Y E R L G F G I V F

Nae I KspI Avr II
CGGGACGCCGCTGGATGATCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATAC
GCGCTGCGGCGGACCTACTAGGAGGTGCGGCCCCTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCTTCTCTGTATTG 400
R D A G W M I L O R G D L M L E F F A H P R G R L T E T R K E T I

KspI
CGGAAGGAACCCGCTATGACGGCAATAAAAGACAGAATAAAACGCACGGTGTTGGGTGTTTGTTCATAAACCGGGGTTGGTCCCAGGGCTGGCA
GCCCTCCTTGGGCGGATACTGCCGTTATTTTCTGTCTTATTTTGCCTGCCACAACCCAGCAAAACAGTATTTGCGCCCCAAGCCAGGGTCCGACCGT 410
P E G T R A M T A I K R O N K T H G V G S F V H K R G V R S Q G W H

CTCTGTCGATACCCACCGAGACCCATTGGGGCCAAACGCCCGCGTTTCTTCTTTTCCCCACCCCAAGTTTCGGGTGAAGGCCAGGGCTC
GAGACAGCTATGGGTGGCTCTGGGGTAACCCGTTATGCGGGCGCAAGAAGGAAAAGGGTGGGGTGGGGGTTCAAGCCCACTTCCGGGTCCCGAG 420
S V D T P P R P H W G O Y A R V S S F S P P H P P S S G E G P G L

AlwI OxaI PraI PraI
GCAGCCAACGTCGGGGCGGCGAGGCCCTGCCATAGCCTCAGGTACTCATATATACTTTAGATTGATTAAACATTCATTTTAATTTAAAGGATCTAGG
CGTCGGTTGCAGCCCCGCGTCCGGGACGGTATCGGAGTCCAATGAGTATATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCC 430
A A N V G A A G P A I A S G Y S Y I L . I D L K L H F . F K R I .

BspH I
TGAAGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCGCTAGAAAAGATCAAAAGATCTTTTGT
ACTTCTAGGAAAACTATTAGAGTACTGGTTTTAGGGAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCTAGAAGAAC 440
V K I L F D N L M T K I P . R E F S F H . A S D P V E K I K G S S .

AGATCCTTTTCTGCGGTAATCTGCTGCTTCCAAACAAAAAACACCGCTACCAGCGGTGGTTTGTTCGCCGATCAAGAGCTACCAACTCTTTT
TCTAGGAAAAAAGACGCGCATTAGACGACGAACCTTTGTTTTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCTCGATGGTTGAGAAAAA 450
D P F F L R V I C C L O T K K P P L P A V V C L P D Q E L P T L F

BsrI
CCGAAGGTAACGGCTTCAGCAGAGCGGAGATACCAATATGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGGCTA
GGCTTCCATTGACCGAAGTCGCTCGCGCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTTCTTGAGACATCGTGCGGAT 460
P K V T G F S R A Q I P N T V L L V . P . L G H H F K N S V A P P

Tuesday, 18 November 1997 10:34
fig 29 pEGFPsac (1 > 5100) Site and Sequence

Page 7

AlwI I
CATACCTCGCTCTGCTAATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCGTGCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTTCAGCACAGAATGGCCCAACCTGAGTTCTGCTATCAATGSCCTATTCCG
T Y L A L L I L L P V A A A S G D K S C L T G L D S R R . L P D K A 4700

ApaI I
GCAGCGGTCGGGCTGAACGGGGGTTCTGTGCACACAGCCAGCTTGGAGCGAACGACCTACACCGAAGTACCTACAGCGTGAGCTATGAGAAAGC
CGTCGCCAGCCGACTTGCCCCCAAGCACGTGTGTCGGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCC
Q R S G . T G G S C T Q P S L E R T T Y T E L R Y L Q R E L . E S 4800

GCCACGCTTCCCAGGGGAGAAAGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCTTGGT
CGGTGCGAAGGGCTTCCCTCTTTCCGCTGTCCATAGGCCATTGCGCTCCAGCCTTGCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGACCA
A T L P E G R K A D R Y P V S G R V G T G E R T R E L P G G N A V 4900

ATCTTTATAGTCTGTCTGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTGTGATGCTCGTCAGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
TAGAAATATCAGGACAGCCAAAGCGGTGGAGACTGAATCGCAGCTAAAAACACTACGAGCAGTCCCCCGCCTCGGATACCTTTTTCGGGTCGTTGCG
Y L Y S P V G F R H L . L E R R F L . C S S G G R S L V K N A S N A 5000

Ava III
Nsi I
GGCCTTTTACGGTTCTTGCCCTTTTGTGGCCTTTTGTTCACATGTTCTTTCTGCGTTATCCCTGATTCTGTGGATAACCGTATTACCGCCATGCAT
CCGGAAAAATGCCAAGGACCGGAAACGACCGGAAACGAGTGTAAGAAAGGACGCAATAGGGGACTAAGACACCTATTGGCATAATGGCGGTACGTA
A F L R F L A F C V P F A H M F F P A L S P D S V D N R I T A M H 5100

Tuesday, 18 November 1997 10:34

fig 30 pEGFP72 (1 > 9697) Site and Sequence

Enzymes : 72 of 146 enzymes (Filtered)

Settings: Linear, Certain Sites Only, Standard Genetic Code

Page 1

Page 8

16p

Bgl I

100

L L I V I N Y G V I S S . P I Y G V P R Y I T Y G K V P A V L T

Aat II

200

A Q R P P P I D V N N D V C S H S N A N R D F P L T S M G G V F T V

Bgl I

Nde I

Aat II

Bgl I

300

N C P L G S T S S V S Y A K Y A P Y . R Q . R . M A R L A L C P V

SnaB I

Nco I

400

H D L M G L S Y L A V H L R I S H R Y Y H G D A V L A V H Q V A V

Aat II

500

I A V . L T G I S K S P P H . R Q V E F V L A P K S T G L S K M S .

Nhe I

Eco47

600

Q L R P I D A N G R . A C T V G G L Y K Q S V F S E P S D P L A L

Nco I

700

P V A T M V S K G E E L F T G V V P I L V E L D G D V N G H K F S

eGFP.C.e.unc53

800

V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W P T L V T

Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 1

CACCC TGACCTACGGCGTGCAGTGC TTCAGCCGCTACCCCGACCATGAAGCAGCAGCTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
GTGGGACTGGATGCCGCACGTACGAAGTCGGCGATGGGGCTGGTGACTTCGTCTGTCTGAAGAAGTTCAGGCGGTACGGGCTTCCGATTCAGGTCTC
900

eGFP.C.e.unc53
T L T Y G V Q C F S R Y P D H M K Q H D F F K S A M P E G Y V Q E

KspI
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGGAGGTGAAGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCG
GCGTGGTAGAAGAAGTTCCTGCTGCCGTTGATGTTCTGGGCGGGCTCCACTTCAAGCTCCCGCTGTGGGACCACTTGGCGTAGCTCGACTTCCCGTAGC
1000

eGFP.C.e.unc53
R T I F F K D D G N Y K T R A E V K F E G D T L V N R I E L K G I

ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAAGCCACAACGCTATATCATGGCCGACAAGCAGAAGAAGCGGCATCAA
TGAAGTTCCTCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGATGTTGTCGGTGTGCAGATATAGTACC GGCTGTTCTGCTCTTCTGCCGTAGTT
1100

eGFP.C.e.unc53
D F K E D G N I L G H K L E Y N Y N S H N V Y I M A D K Q K N G I A

GGTGAAC TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGCTGCTG
CCACTTGAAGTTC TAGGCGGTGTTG TAGCTCCTGCCGTCGCAGTCGAGCGGCTGGTGATGGTCTCTTGTGGGGTAGCCGCTGCCGGGGCACGACGAC
1200

eGFP.C.e.unc53
V N F K I R H N I E D G S V Q L A D H Y Q Q N T P I G D G P V L L

CCCGACAACCACTACCTAGACACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCACATGGTCC TGCTGGAGTTCGTGACC3CCGCCGGGA
GGGCTGTTGGTGATGGACTCGTGGGTCAGGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTAACAGGACGACCTCAAGCACTGGCGGGCGCCCT
1300

eGFP.C.e.unc53
P D N H Y L S T O S A L S K D P N E K R D H M V L L E F V T A A G

BspM II Bgl II
TCACTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTACGTCAAATGTAGAATTGATACCAATCTACACGGATTGGGCCAATCGGCACCTTTC
AGTGAGAGCCGTACCTGCTCGACATGTTAGGCCCTGAGTCTAGATGCAGTTTACATCTTAACATATGGTTAGATGTGCC TAACCCGGTTAGCGGTGGAAAG
1400

eGFP.C.e.unc53
I T L G M D E L Y K S G L R S T S N V E L I P I Y T D V A N R H L S

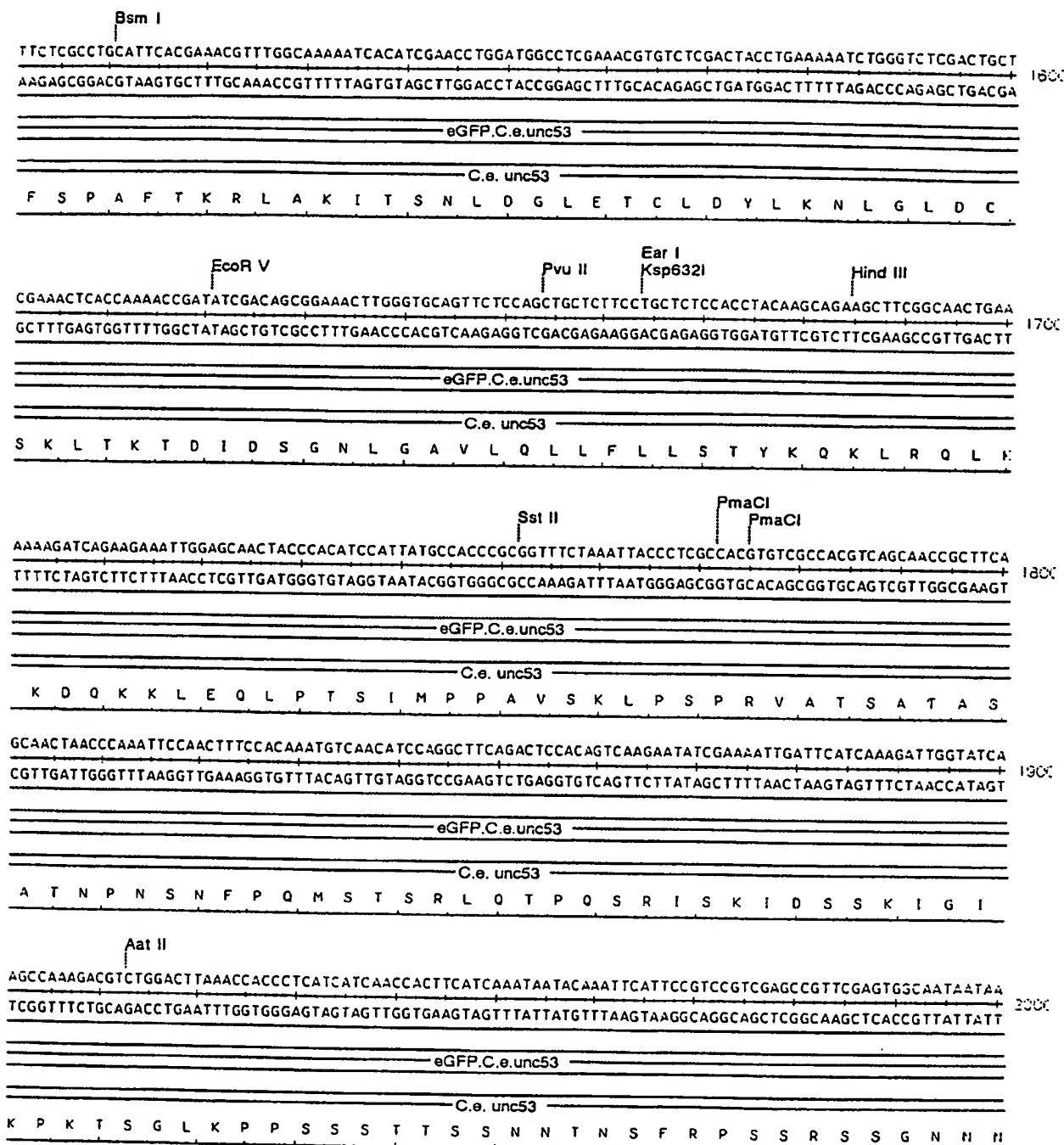
C.e.unc53
GAAGGGCAGCTTATCAAAGTCGATTAGGGATATTTCCAATGATTTTCGGGACTATCGACTGGTTTCTCAGCTTATTAATGTGATCGTTCCGATCAACGAA
CTTCCCGTCGAATAGTTTCAAGTAATCCCTATAAAGGTTACTAAAAGCGCTGATAGCTGACCAAAGAGTCGAATAATTACACTAGCAAGGCTAGTTGCTT
1500

eGFP.C.e.unc53
K G S L S K S I R D I S H D F R D Y R L V S Q L I N V I V P I N E

Nru I EcoR I

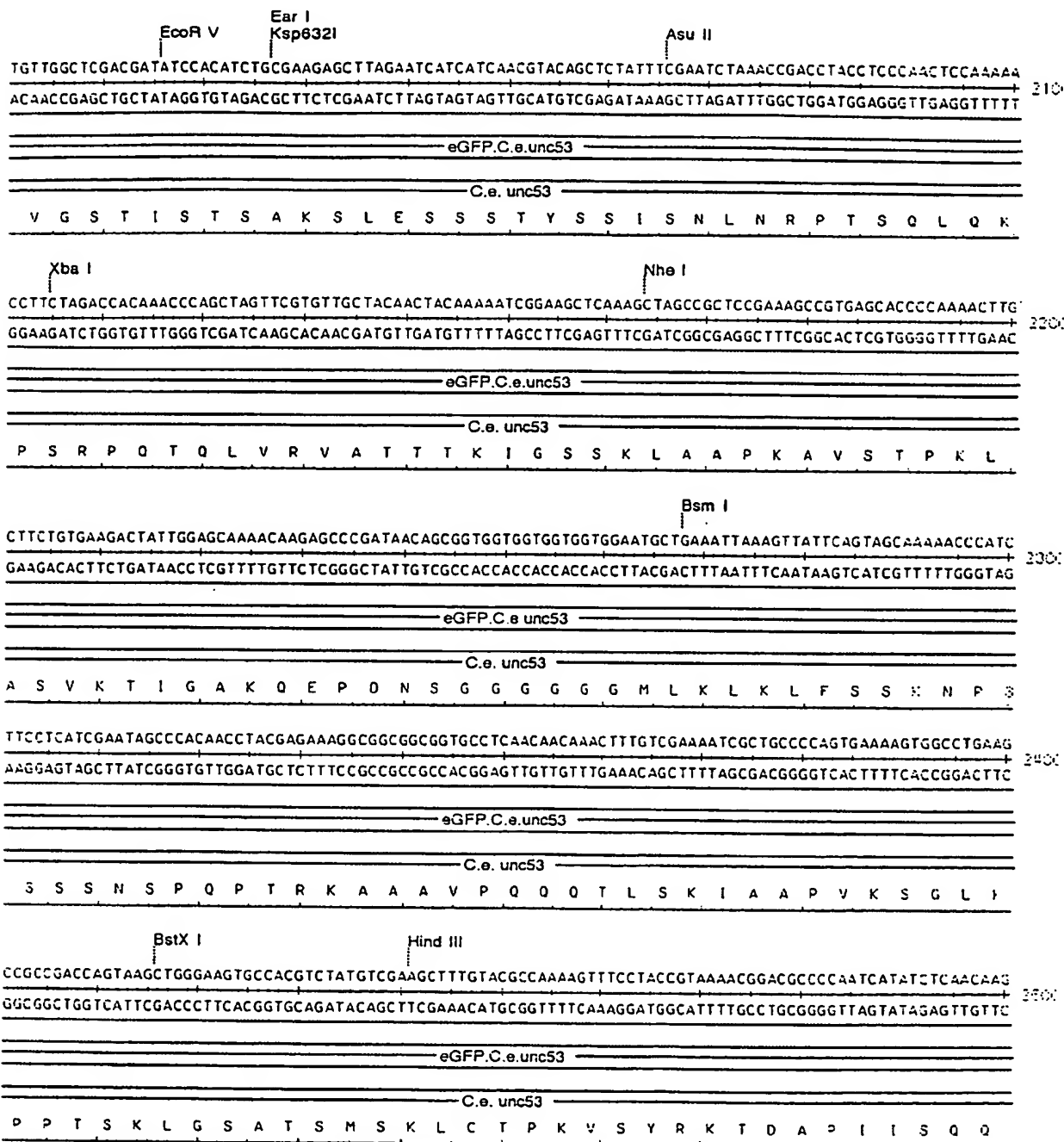
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 3



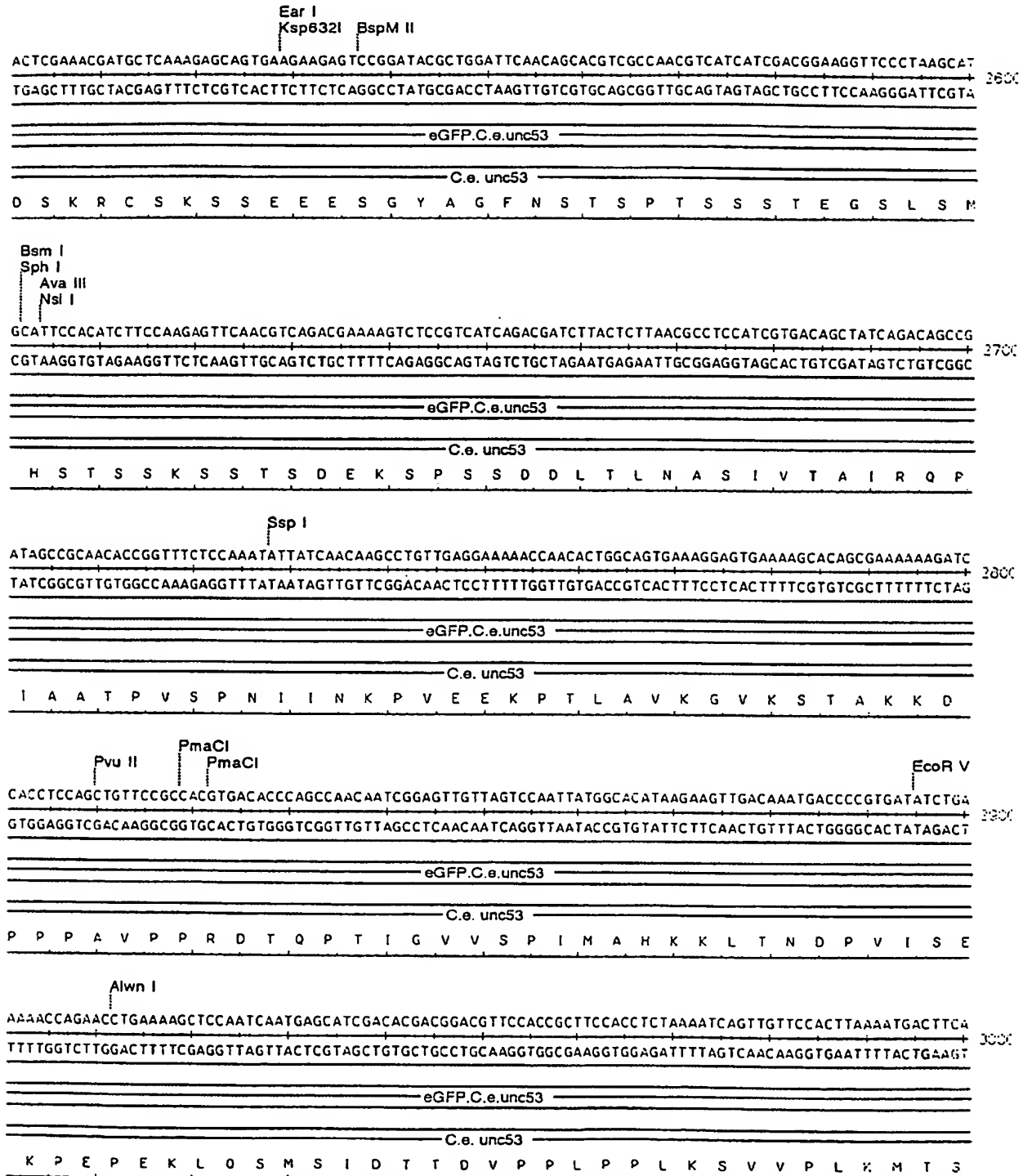
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 4



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 6



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

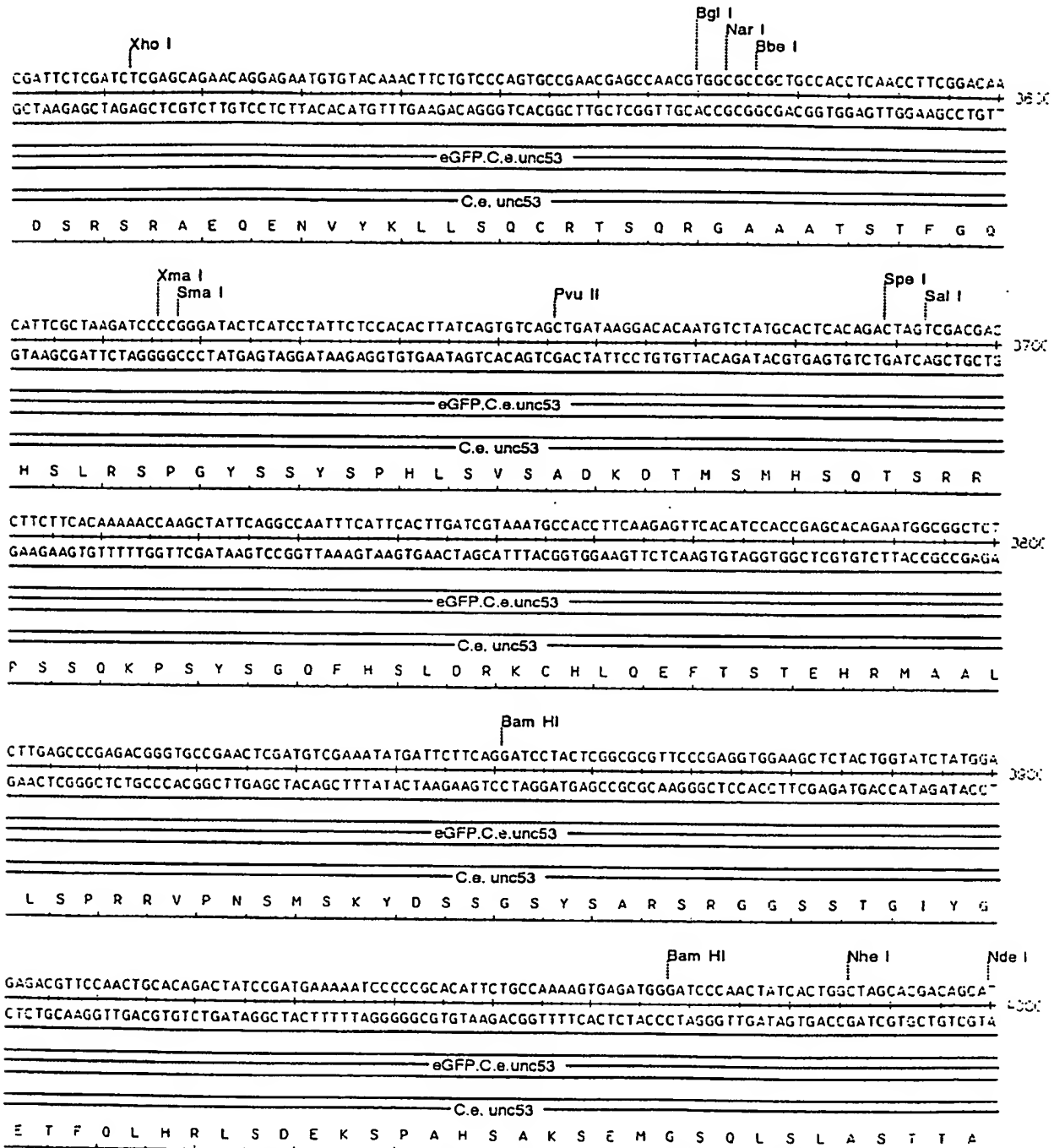
Page 4

Spl
Spl

ATCCGACAACCAACGTACGATGTTCTTCTAAAACAAGGAAAAATCACATCGCCTGTCAAGTCGTTTGGATATGAGCAGTCGTCGCGTCTGAAGACT 310
TAGGCTGTTGGTGGTTGCATGCTACAAGAAGATTTTGTTCCTTTTAGTGTAGCGGACAGTTTACGAAACCATACTCGTCAGCAGGCGCAGACTTCTGA
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
I R Q P P T Y D V L L K Q G K I T S P V K S F G Y E Q S S A S E D
CCATTGTGGCTCATGCGTCGGCTCAGGTGACTCCGCCGACAAAACTTCTGGTAATCATTGCTGGAGAGAAGGATGGGAAAGAATAAGACATCAGAATC 320
GGTAACACCGAGTACGCAGCCGAGTCCACTGAGGCGGCTGTTTTGAAGACCATTAGTAAGCGACCTCTCTTCCTACCCTTTCTTATTCTGTAGCTTAG
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S I V A H A S A Q V T P P T K T S G N H S L E R R M G K N K T S E S
CAGCGGTACACCTCTGACGCCGGTGTGCGATGTGCGCAAAATGAGGGAGAAGCTGAAAGAATACGATGACATGACTCGTCGAGCACAGAACGGCTAT 330
GTCGCCGATGTGGAGACTGCGGCCACAACGCTACACGCGGTTTTACTCCCTCTTCGACTTCTTATGCTACTGTACTGAGCAGCTCGTGCTTTGCCGATA
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S G Y T S D A G V A M C A K M R E K L K E Y D D M T R R A Q N G Y
Asu II Sst I BspM II
CCTGACAACCTCGAAGACAGTTCCTCCTTGTGCTGGAATATCCGATAACAACGAGCTCGACGACATATCCACGGACGATTTGTCCGGAGTAGACATGG 340
GGACTGTTGAAGCTTCTGTCAAGGAGGAACAGCAGACCTTATAGGCTATTGTTGCTCGAGCTGCTGTATAGGTGCCTGCTAAACAGGCCATCTGTATCC
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
P D N F E D S S S L S S G I S D N N E L D D I S T D D L S G V D H
CAACAGTCGCCTCCAAACATAGCGACTATTCCCACTTTGTTTCGCCATCCACGCTCTCTTCTCAAAGCCCGAGTCCCAGTCGGTCTCCACATCAGT 350
GTTGTCAGCGGAGGTTTGTATCGCTGATAAGGGTGAACAAAGCGGTAGGGTGCAGAAGAAGGAGTTTCGGGGCTCAGGGGTCAGCCAGGAGGTTAGTCA
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
A T V A S K H S D Y S H F V R H P T S S S S K P R V P S R S S T S V

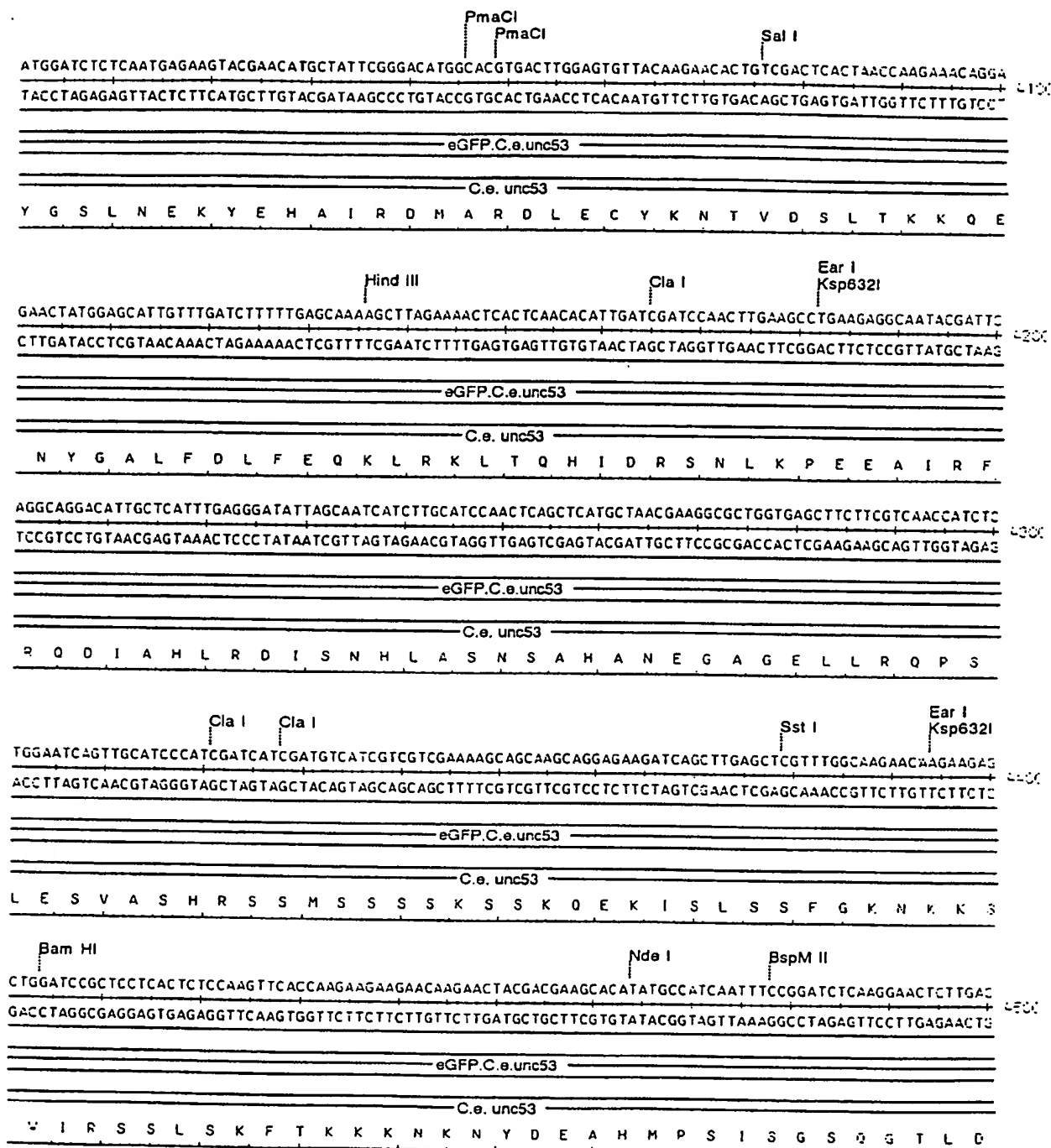
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 7



Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 8



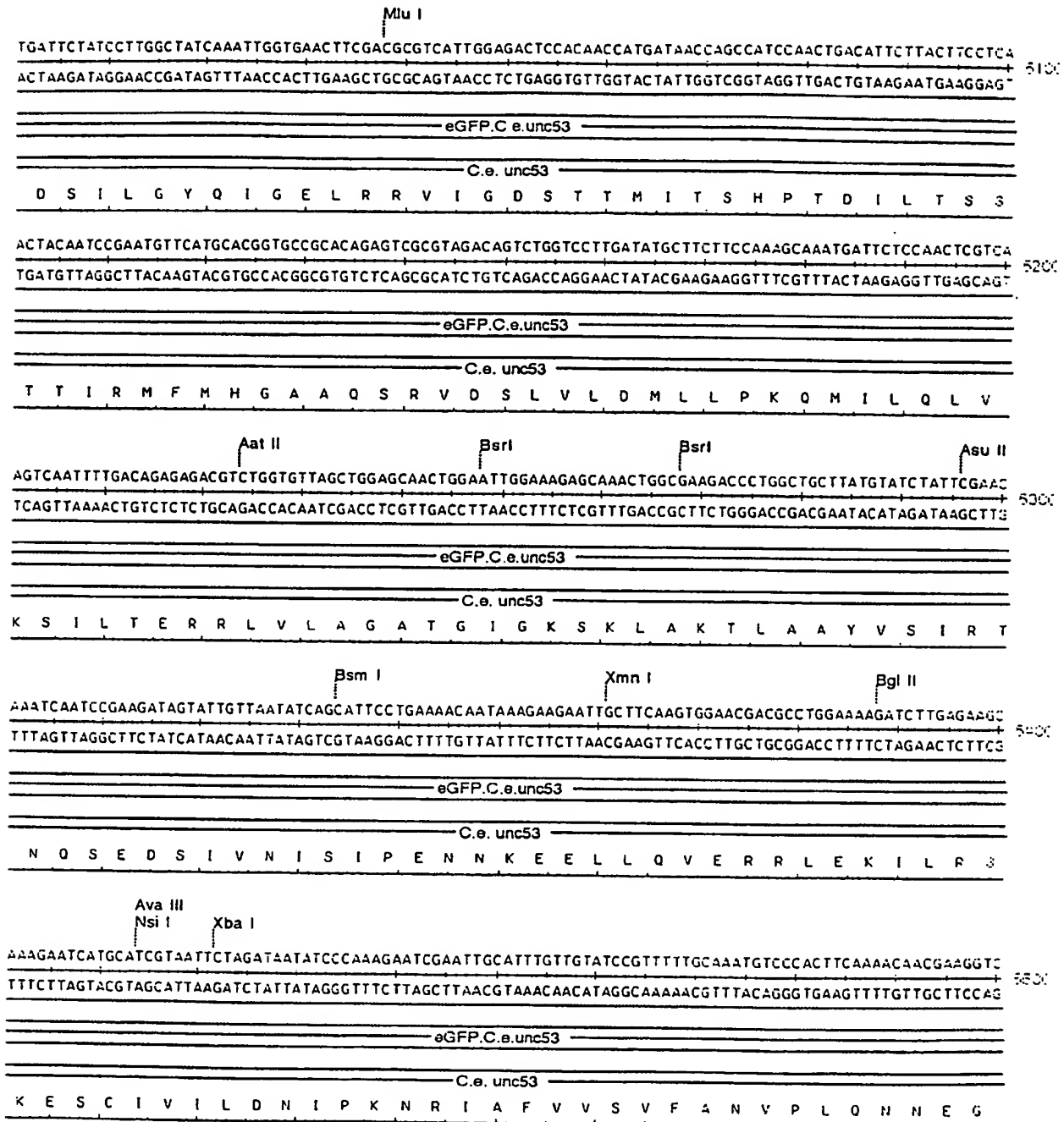
Tuesday, 18 November 1997 10:34
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 9

Sst I Apa I
AACATTGATGTGATTGAGTTGAAGCAAGAGCTCAAAGAACGCGATAGTGCACCTTTACGAAGTCCGCTTGACAATCTGGATCGTGCCCGC3AAGTTGATG
TTGTAAC TACACTAACTCAACTTCGTTCTCGAGTTTCTTGCGCTATCACGTGAAATGCTTCAGGCGGAAC TGT TAGACCTAGCACGGGCGCTTCAACTAC 4600
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
N I D V I E L K O E L K E R D S A L Y E V R L D N L D R A R E V D
TTCTGAGGGAGACAGTGAACAAGTTGAAAACCGAGAACAAGCAATTAAGAAAGAGTGGACAACTCACCAACGGTCCAGCCACTCGTGCTTCTTCCCG
AAGACTCCCTCTGTCACTTGTTCAACTTTTGCTCTTGTTCGTTAATTCTTTCTTCCACCTGTTTGAGTGGTTGCCAGGTCGGTGAGCAC3AAGAGGGG 4700
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
V L R E T V N K L K T E N K Q L K K E V D K L T N G P A T R A S S R
Ksp I Bsr I Asu II
CGCCTCAATTCAGTTATCTACGACGATGAGCATGTCATGATGCAGCGTGTAGCAGTACATCAGCTAGTCAATCTTCGAAACGATCCTCTGGCTGCAAC
GCGGAGTTAAGGTCAATAGATGCTGCTACTCGTACAGATACTACGTCGCACATCGTCATGTAGTCGATCAGTTAGAAGCTTTGCTAGGAGACCGACGTTG 4800
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
A S I P V I Y D D E H V Y D A A C S S T S A S Q S S K R S S G C H
Pvu I Hpa I EcoR V
TCAATCAAGGTTACTGTAAACGTGGACATCGCTGGAGAAATCAGTTTCGATCGTTAACCCGGACAAAGAGATAATCGTAGGATATCTTGCCATGTCAACCA
AGTTAGTTCCAATGACATTTGCACCTGTAGCGACCTCTTTAGTCAAGCTAGCAATTGGGCTGTTTCTCTATTAGCATCCTATAGAACGGTACAGTTGG 4900
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S I K V T V N V D I A G E I S S I V N P D K E I I V G Y L A M S T
Cla I
GTCAGTCATGCTGGAAAGACATTGATGTTTCTATTCTAGGACTATTTGAAGTCTACCTATCCAGAATTGATGTGGAGCATCAACTTGGAAATCGATGCTG
CAGTCAGTACGACCTTTCTGTAACTACAAGATAAGATCCTGATAAATTTCAGATGGATAGGTCTTAAC TACACCTCGTAGTTGAACCTTAGCTACGAGC 5000
-----eGFP.C.e.unc53-----
-----C.e. unc53-----
S O S C V K D I D V S I L G L F E V Y L S R I D V E H Q L G I D A R

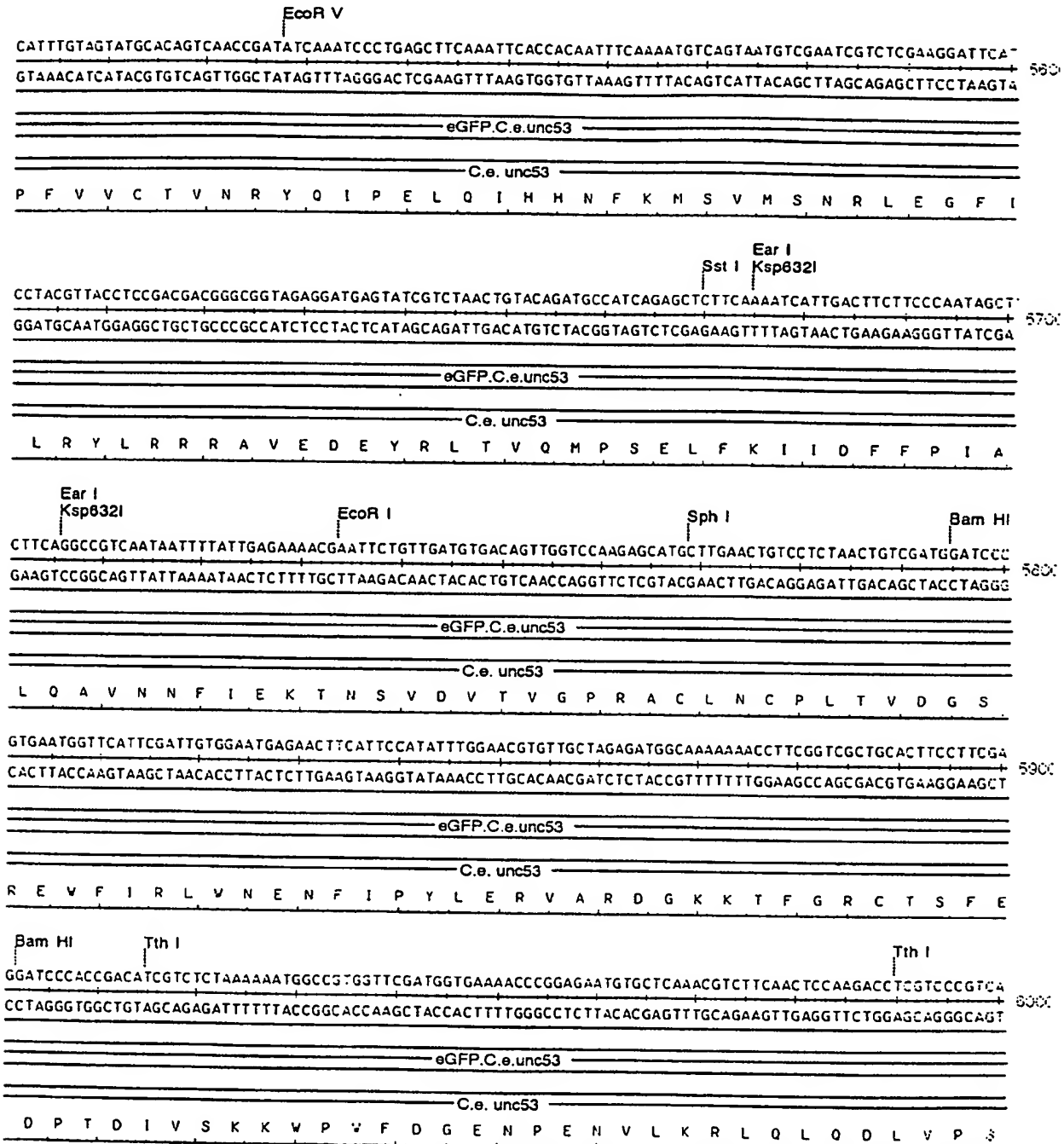
Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 10

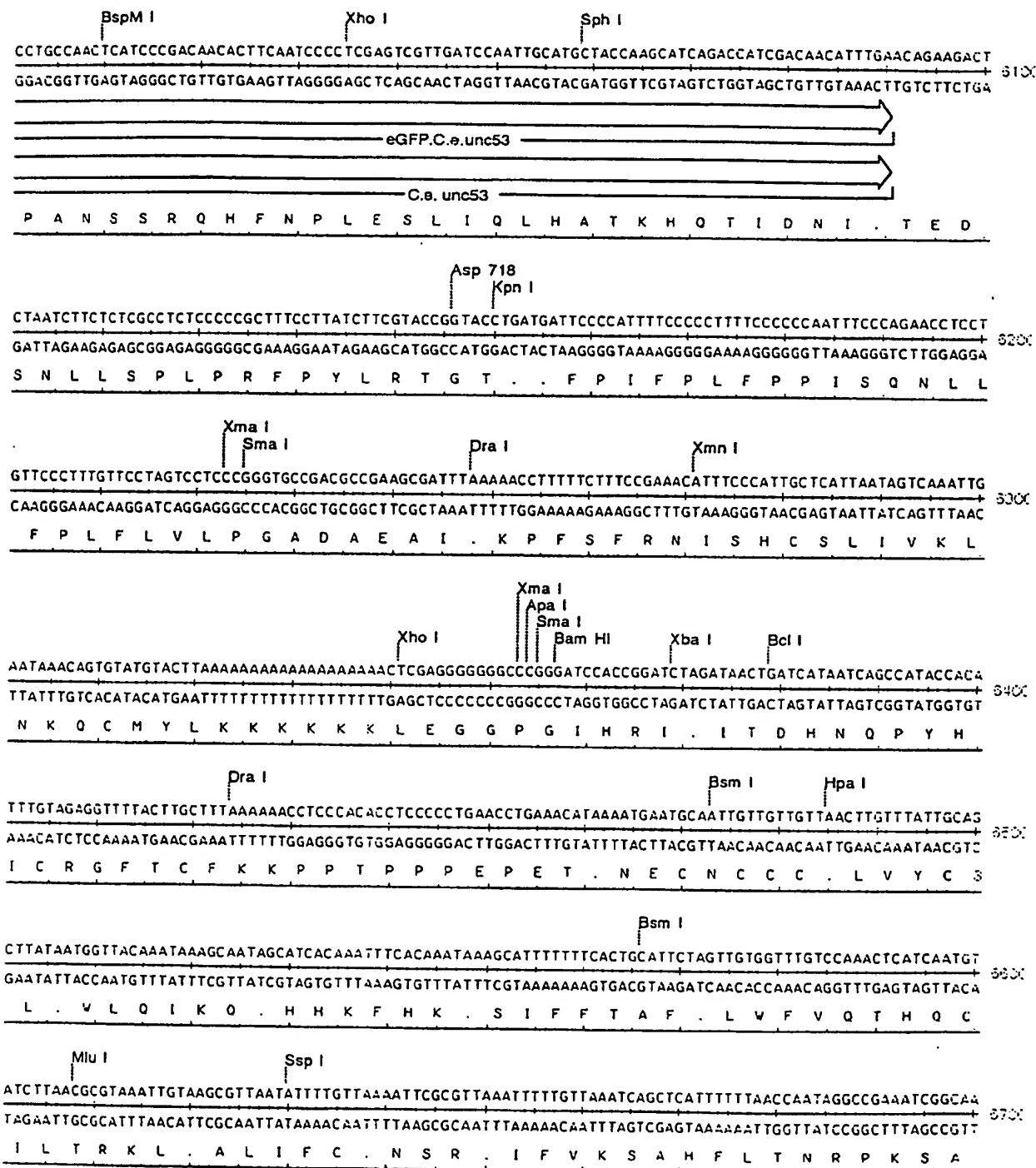


Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 11



Page 12



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 13

Bsr I

AATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAGAAGCTGGACTCCAACGTCAAA
TTAGGGAATATTAGTTTCTTATCTGGCTCTATCCCAACTCACAAAGSTCAAACCTTGTCTCAGGTGATAATTTCTTGACCTGAGGTTGCAGTTT
K S L I N Q K N R P R . G . V L F Q F G T R V H Y . R T V T P T S I 5800

Dra III

GGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACC
CCCCTTTTGGCAGATAGTCCCGCTACCGGGTGATGCACCTTGGTAGTGGGATTAGTTCAAAAACCCAGCTCCACGGCATTTCGTGATTAGCCTTGG
G E K P S I R A M A H Y V N H H P N Q V F W G R G A V K H . I G T 5900

Nae I

CTAAGGGAGCCCCGATTAGAGCTTGACGGGGAAGCGCGCAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGC
GATTTCCCTCGGGGGCTAAATCTCGAACTGCCCCTTTCGGCCGCTTGCACCGCTCTTTCTTCCCTTCTTCGCTTCTCGCCCGCGATCCCGGACCG
L K G A P D L E L D G E S R R T V R E R K G R K R K E R A L G R V 7000

Ksp I

AAGTGAGCGGTACGCTGCGGTAACCACACACCCGCGCGCTTAATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
TTCACATCGCCAGTGCGACGCGCATTGGTGGTGTGGCGGCGCGAATTACGCGCGCATGTCGCCGCGAGTCCACCGTGAAAAGCCCCTTTACACGCGCCT
Q V . R S R C A . P P H P P R L M R R Y R A R Q V A L F G E M C A E 7100

Bsp H I

Ssp I

Ear I
Ksp 6321

ACCCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCGTGATAAATGCTTCAATAATATTGAAAAGGAAGAGTC
TGGGATAAACAAATAAAAGATTATGTAAGTTTATACATAGGCGAGTACTCTGTTATTGGGACTATTTACGAAGTTATTATAACTTTTCTTCTCAG
P L F V Y F S K Y I O I C I R S . D N N P D K C F N N I E K G R V 7200

Oxa N I

Pvu II

Sph I
Ava III
Nsi I

CTGAGGCGGAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAGAAGTATGCAAAGCATGCATCTCAAT
GACTCCGCTTTCTTGGTCGACACCTTACACACAGTCAATCCACACCTTTCAGGGGTCCGAGGGGTGCTCCGCTTTCATACGTTTCGTACGTAGAGTTA
L R R K E P A V E C V S V R V W K V P R L P S R Q K Y A K H A S Q 7300

Sph I
Ava III
Nsi I

TAGTCAGCAACCAAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAA
ATCAGTCGTTGGTCCACACCTTTTCAGGGGTCCGAGGGGTGCTCCGCTTTCATACGTTTCGTACGTAGAGTTAATCAGTCGTTGGTATCAGGGCGGGGATT
L V S N Q V W K V P R L P S R Q K Y A K H A S Q L V S N H S P A P N 7400

Bsr I

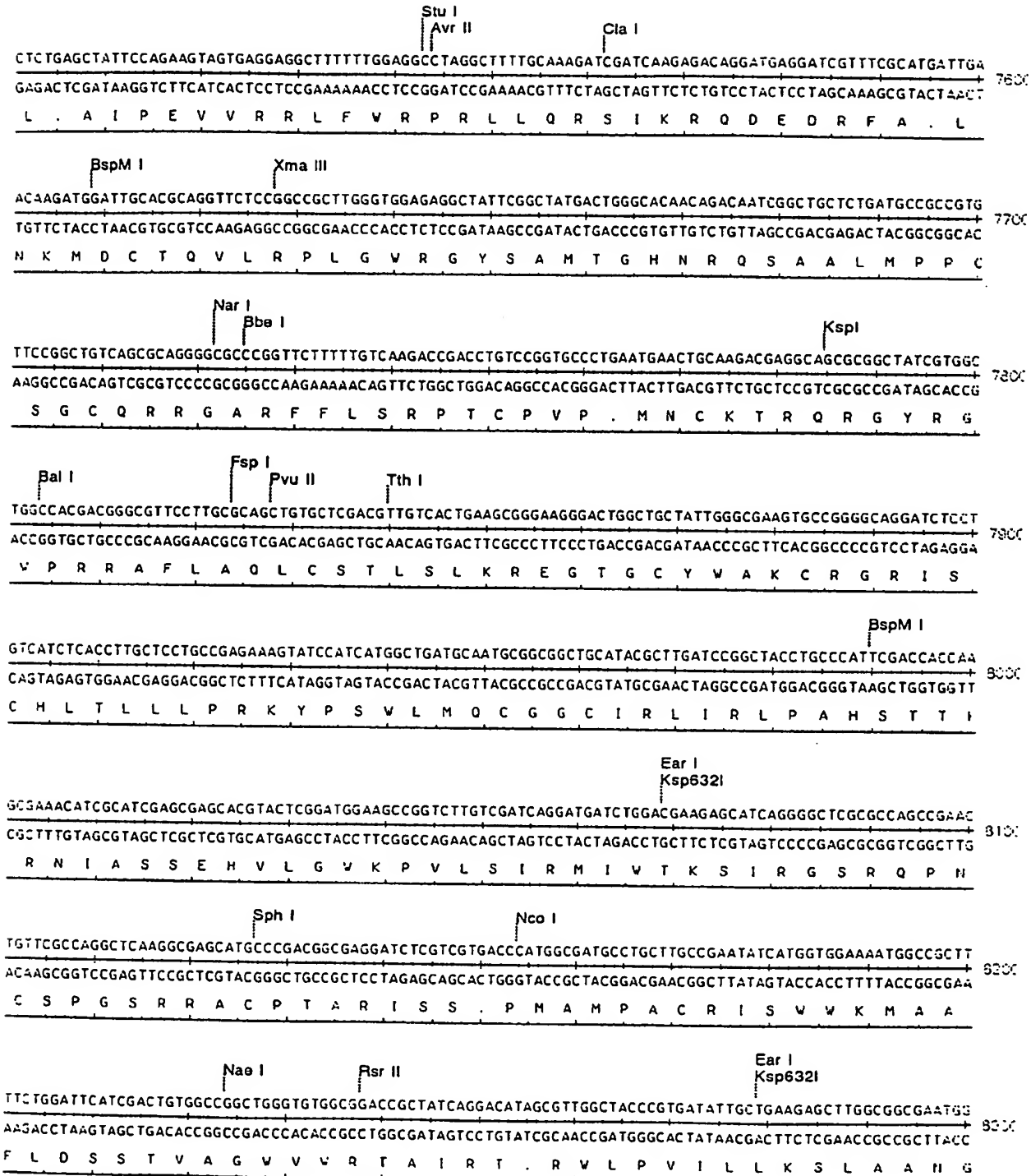
Nco I

Bgl I
Sfi I

CTCCGCCATCCCGCCCTAACTCCGCCAGTTCGCCCATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGC
GAGGCGGGTAGGGCGGGGATTGAGGCGGGTCAAGGCGGGTAAAGAGCGGGGTACCGACTGATTAAAAAAATAAATACGTCTCCGGCTCCGGCGGAGCCG
S A H P A P N S A Q F R P F S A P V L T N F F Y L C R G R G R L G 7500

Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 14



Tuesday, 18 November 1997 10:35
fig 30 pEGFP72 (1 > 9697) Site and Sequence

Page 14

GCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTGCGAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGSS
CGACTGGCGAAGGAGCAGGAAATGCCATAGCGGCGAGGGCTAAGCGTCGCGTAGCGGAAGATAGCGGAAGAAGTGTCAAGAAGACTCGCCCTGAGACCC 940
L T A S S C F T V S P L P I R S A S P S I A F L T S S S E R D S G

Asu II BspM I
GTTGAAATGACCGACCAAGCGACGCCAACCTGCCATCACGAGATTTGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCGG 950
CAAGCTTTACTGGCTGGTTGCTGCGGGTTGGACGGTAGTGCTCTAAAGCTAAGGTGGCGGCGGAAGATACTTTCCAACCCGAAGCCTTAGCAAAAGGCC
V R N D R P S D A Q P A I T R F R F H R R L L . K V G L R N R F P

Nae I KspI Avr II
GACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCG 960
CTGCGGCGGACCTACTAGGAGGTGCGGCCCTTAGAGTACGACCTCAAGAAGCGGGTGGGATCCCCCTCCGATTGACTTTGTGCCTTCTCTGTTATGGCC
G R R L D D P P A R G S H A G V L R P P . G E A N . N T E G D N T G

KspI
AAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAAAACGCACGGTGTGGGTCGTTTGTTCATAAACCGGGGTTGCGTCCAGGGCTGGCACTC 970
TTCTTGGGCGGATACGCGGTTATTTTCTGTCTTATTTTGGTGCCACAACCCAGCAACAAGTATTTGCGCCCCAAGCCAGGGTCCGACCGTGAG
R N P R Y D G N K K T E . N A R C V V V C S . T R G S V P G L A L

TGTCGATACCCACCGAGACCCCATTTGGGGCCAATACGCCCGGTTTCTTCTTTTCCCCACCCACCCCAAGTTCGGGTGAAGGCCAGGGCTCGCA 980
ACAGCTATGGGGTGGCTCTGGGGTAACCCCGGTTATGCGGGCGCAAGAAGGAAAGGGGTGGGGTGGGGGGTTCAAGCCCACTTCCGGGTCGCGAGCGT
C R Y P T E T P L G P I R P R F F L F P T P P P K F G . R P R A R

AlwI OxaN I Dra I Dra I
GCCAAGCTCGGGGCGGCGAGCCCTGCCATAGCCTCAGGTACTCATATATACTTTAGATTGATTTAAACCTTCATTTTTAATTTAAAGGATCTAGGTGA 990
CGGTTGACGCCCCCGCTCCGGGACGGTATCGGAGTCCAATGAGTATATGAAATCTAACTAAATTTGAAGTAAAAATTAATTTTCTAGATCCACT
S Q R R G G R P C H S L R L L I Y T L D . F K T S F L I . K D L G E

BspH I
AGATCCTTTTGTATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGA 1000
TCTAGGAAAAATATTAGAGTACTGGTTTATGGGAATTGCACTCAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTCTAGTTTCTAGAGAAGAACTCT
D P F . . S H D Q N P L T . V F V P L S V R P R R K D Q R I F L R

TCCTTTTCTGCGGTAATCTGCTGCTTGCAACAAAAAACCCCGCTACCAGCGGTGGTTTGTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCG 1010
AGGAAAAAAGACGCGCATTAGACGACGAACGTTGTTTTTGGTGGCGATGGTCGCCACCAACAAACGGCCTAGTTCGATGGTTGAGAAAAAGGC
S F F S A R N L L L A N K K T T A T S G G L F A G S R A T N S F S

BsrI
AAGGTAACCTGGCTTCAGCAGAGCGCAGATACCAAACTACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT 1020
TTCCATTGACCGAAGTCGTCTCGGCTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAAGTCTTTGAGACATCGTGGCGGATGT
E G N V L Q Q S A D T K Y C P S S V A V V R P P L Q E L C S T A Y I



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, 5/10, 15/85, C07K 14/435, 16/18, A61K 38/17, 49/00, C12Q 1/02, G01N 33/53	A3	(11) International Publication Number: WO 98/24810 (43) International Publication Date: 11 June 1998 (11.06.98)
(21) International Application Number: PCT/EP97/06956 (22) International Filing Date: 3 December 1997 (03.12.97) (30) Priority Data: 9625283.8 4 December 1996 (04.12.96) GB (71) Applicant (for all designated States except US): JANSSEN PHARMACEUTICA N.V. [BE/BE]; Turnhoutseweg 30, B-2340 Beerse (BE). (72) Inventors; and (75) Inventors/Applicants (for US only): PLATTEEUW, Christ, Jules [BE/BE]; Evergemsesteenweg 17, B-9032 Wondel- gem (BE). BUESA ARJOL, Carlos, Manuel [ES/ES]; Trav- essera de les Corts, 171/702a, E-08028 Barcelona (ES). DERAEMYAEKER, Marc [BE/BE]; Janssen Pharmaceu- tica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). VER- HASSELT, Peter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). PUJOL, Nathalie, Jeanne, Raymonde [FR/BE]; 213, avenue du Père Soulas, F-34000 Montpellier (FR). MAERTENS, Luc, Jacques, Si- mon [BE/BE]; Vier Uitersten 26, B-8200 Brugge (BE). LUYTEN, Walter [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). GEERTS, Hugo [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30,		B-2340 Beerse (BE). VANDEKERCKHOVE, Joel, Ste- faan [BE/BE]; Rode Boukendreef 27, B-8210 Loppem (BE). GEYSEN, Johan [BE/BE]; Janssen Pharmaceutica N.V., Turnhoutseweg 30, B-2340 Beerse (BE). BOGAERT, Thierry, André, Olivier, Eddy [BE/BE]; Wolvendreef 26g, B-8500 Kortrijk (BE). (74) Agent: BALDOCK, Sharon, Claire; Boulton Wade Tennant, 27 Furnival Street, London EC4A 1PQ (GB). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims</i> <i>and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 21 January 1999 (21.01.99)
(54) Title: VERTEBRATE HOMOLOGUES OF UNC-53 PROTEIN OF C. ELEGANS (57) Abstract Vertebrate protein homologues of UNC-53 protein of C. elegans and nucleic acid sequences coding for said homologues or functional equivalents thereof are identified. The nucleic acid sequences in an appropriate vector are used to transfect or transform cells, tissues or organisms useful in identifying inhibitors or enhancers of the vertebrate homologue, or further proteins involved in the signal transduction pathway of which said vertebrate homologue is a component. Any of said inhibitors or enhancers identified can be included in a pharmaceutical composition or in the preparation of a medicament for treating conditions such as neurological diseases, acute traumatic injuries and to promote neuronal regeneration and inhibit metastasis or loss of contact inhibition.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

Internat. Application No
PCT/EP 97/06956

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C12N5/10 C12N15/85 C07K14/435 C07K16/18
A61K38/17 A61K49/00 C12Q1/02 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	WO 96 38555 A (BOGAERT THIERRY (BE); STRINGHAM EVE (CA); VANDEKERCKHOVE JOEL (BE)) 5 December 1996	1, 2, 24-26, 28, 30-36, 38, 40, 42-51, 66, 68, 70, 72, 78, 79, 83-87, 98-105, 107, 112
P, Y	see page 2, line 18 - page 20, line 26	3-23, 27, 29, 37, 39, 41, 58, 59, 67, 69, 74, 80-82,
	-/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

2 October 1998

Date of mailing of the international search report

27. 11. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Donath, C

INTERNATIONAL SEARCH REPORT

Internat'l Application No

PCT/EP 97/06956

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	<p>see figure 5</p> <p>see Sequence Listing: SEQ ID NO.: 1 and 4</p> <p style="text-align: center;">---</p>	<p>88, 115-118, 124,125</p>
Y	<p>HEKIMI, S. AND KERSHAW, D.: "Axonal guidance defects in a Caenorhabditis elegans mutant reveal cell-extrinsic determinants of neuronal morphology" THE JOURNAL OF NEUROSCIENCE, vol. 13, no. 10, October 1993, pages 4254-4271, XP000612286</p> <p>see page 4254, left-hand column, paragraph 1 - page 4255, right-hand column, paragraph 2</p> <p>see page 4267, right-hand column, paragraph 1 - page 4271, left-hand column, paragraph 2</p> <p style="text-align: center;">---</p>	<p>3-23,27, 29,37, 39,41, 58,59, 67,69, 74, 80-82, 88, 115-118, 124,125</p>
A	<p style="text-align: center;">---</p> <p>STERN, M.J. ET AL.: "The human GRB2 and Drosophila Drk genes can functionally replace the Caenorhabditis elegans cell signaling gene sem-5" MOLECULAR BIOLOGY OF THE CELL, vol. 4, no. 11, November 1993, pages 1175-1188, XP002079466</p> <p>cited in the application</p> <p>see the whole document</p> <p style="text-align: center;">-----</p>	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 97/06956

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Although claims 49-51, 58, 59, 107, 115 and 118 are directed to a diagnostic method practised on the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 52-57, 60-65, 71, 73, 75-77, 89-94, 108-111, 113, 114
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 52-57,60-65,71,73,75-77,89-94,108-111,113,114

Claims 52-57, 60-65, 75-77, 91-94, 108-111,113 and 114 concern a compound, pharmaceutical composition, a nucleic acid sequence or the use of a compound. These compounds or compositions, however, are only defined by the method which can be used in order to identify these compounds or compositions as enhancers or inhibitors of regulation of cell shape, cell growth or motility or of the direction of cell migration or of the signal transduction pathway. Since it is completely unclear which kind of substances will be identified by the respective method and since in the specification no concrete examples for these kind of substances are given, the scope of said claims is totally ambiguous and undefined. Moreover, it cannot be excluded that even substances known in the art may be recognized as an enhancing or inhibiting compound by the respective used method.

The same applies to claims 71, 73, 89 and 90 concerning methods comprising proteins which are only defined by a reference to another method which was used to identify said proteins.

Information on patent family members

PCT/EP 97/06956

Form PCT/ISA/210 (patent family annex) (July 1992)